

# Analytic Continuation Of Chern-Simons Theory

Edward Witten

*School of Natural Sciences, Institute for Advanced Study  
Einstein Drive, Princeton, NJ 08540 USA*

## Abstract

The title of this article refers to analytic continuation of three-dimensional Chern-Simons gauge theory away from integer values of the usual coupling parameter  $k$ , to explore questions such as the volume conjecture, or analytic continuation of three-dimensional quantum gravity (to the extent that it can be described by gauge theory) from Lorentzian to Euclidean signature. Such analytic continuation can be carried out by generalizing the usual integration cycle of the Feynman path integral. Morse theory or Picard-Lefschetz theory gives a natural framework for describing the appropriate integration cycles. An important part of the analysis involves flow equations that turn out to have a surprising four-dimensional symmetry. After developing a general framework, we describe some specific examples (involving the trefoil and figure-eight knots in  $S^3$ ). We also find that the space of possible integration cycles for Chern-Simons theory can be interpreted as the “physical Hilbert space” of a twisted version of  $\mathcal{N} = 4$  super Yang-Mills theory in four dimensions.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Overview</b>	<b>7</b>
2.1	Finite-Dimensional Analogs . . . . .	8
2.2	Role Of Morse Theory . . . . .	10
2.3	Stokes Phenomena . . . . .	11
2.4	Singularities Of The Analytically Continued Theory . . . . .	12
2.5	Analytic Continuation For Compact Gauge Groups . . . . .	12
2.5.1	Downward Flow Again . . . . .	14
2.5.2	Indeterminacy of $\mathcal{C}$ . . . . .	16
2.5.3	Knots . . . . .	16
2.6	Comparing Compact And Complex Lie Groups . . . . .	17
2.6.1	Knots Again . . . . .	18
2.7	Other Real Forms . . . . .	20
2.7.1	Comparison Of Holomorphic Volume Forms . . . . .	21
<b>3</b>	<b>Finite-Dimensional Oscillatory Integrals, Morse Theory, And Steepest Descent</b>	<b>23</b>
3.1	The Airy Function . . . . .	23
3.1.1	Morse Theory . . . . .	24
3.1.2	Crossing A Stokes Ray . . . . .	28
3.1.3	Asymptotic Behavior Of Integrals Over Lefschetz Thimbles . . . . .	30

3.1.4	Sum Of Asymptotic Series . . . . .	30
3.1.5	Expressing The Real Cycle In Terms Of Lefschetz Thimbles . . . . .	31
3.1.6	The Asymptotic Behavior Of The Airy Function . . . . .	32
3.2	Generalization To Dimension $n$ . . . . .	34
3.2.1	Essential Singularities . . . . .	37
3.2.2	A Special Situation . . . . .	39
3.3	Symmetries . . . . .	41
3.3.1	Oscillatory Integrals On More General Manifolds . . . . .	41
3.3.2	Manifolds Of Critical Points . . . . .	42
3.3.3	Group Actions . . . . .	43
3.3.4	More on Group Actions . . . . .	46
3.4	Singularities . . . . .	48
3.4.1	More on the Airy Function . . . . .	48
3.4.2	An Example With $SO(2)$ Symmetry . . . . .	50
3.4.3	One More Case . . . . .	54
3.5	Analytic Continuation In $k$ : A Prototype . . . . .	55
3.5.1	Description By Lefschetz Thimbles . . . . .	58
3.5.2	Singularities At $\rho = \pm 1$ . . . . .	59
<b>4</b>	<b>Application To Chern-Simons Theory</b>	<b>62</b>
4.1	The Flow Equation For Chern-Simons . . . . .	63
4.1.1	The Index . . . . .	66

4.1.2	Applications Of The Flow Equation . . . . .	68
4.2	Knots And Morse Theory . . . . .	71
4.2.1	Classical Equations In The Presence Of A Knot . . . . .	73
4.2.2	More On The Classical Equations . . . . .	74
4.2.3	Analog For Noncompact Lie Groups . . . . .	77
4.2.4	Classification Of Critical Points . . . . .	78
4.2.5	Variation Of The Action . . . . .	80
4.2.6	Branches Of The Logarithm . . . . .	83
4.2.7	Tensoring With A Flat Line Bundle . . . . .	84
<b>5</b>	<b>Analytic Continuation Of The Colored Jones Polynomial</b>	<b>85</b>
5.1	Preliminaries . . . . .	86
5.1.1	Local Models . . . . .	89
5.1.2	Normalization Of The Path Integral . . . . .	91
5.1.3	Orientation Of The Integration Cycle . . . . .	93
5.2	The Trefoil Knot . . . . .	94
5.2.1	Singularities Of The Moduli Space Of Flat Connections . . . . .	95
5.2.2	The Chern-Simons Function . . . . .	99
5.2.3	The Integration Cycle And The Asymptotic Behavior Of The Path Integral . . . . .	100
5.2.4	Stokes Curves . . . . .	103
5.3	The Figure-Eight Knot . . . . .	108

5.3.1	Real $\mathfrak{m}$ . . . . .	109
5.3.2	$\mathfrak{m}$ Of Modulus 1 . . . . .	111
5.3.3	The Real Part Of The Chern-Simons Function . . . . .	113
5.3.4	The Imaginary Part Of $W$ . . . . .	116
5.3.5	The Volume Conjecture For The Figure-Eight Knot . . . . .	117

## 6 Four Or Five-Dimensional Interpretation 121

# 1 Introduction

This paper is devoted to the analytic continuation of three-dimensional Chern-Simons gauge theory as a function of its coupling parameters. One motivation for this study concerns three-dimensional quantum gravity, which plausibly has at least some sort of relationship to three-dimensional Chern-Simons theory with gauge group  $SL(2, \mathbb{C})$ . In that context, Euclidean quantum gravity is obtained by analytic continuation to imaginary values of the coupling parameter  $s$  that is introduced in section 2.<sup>1</sup> However, for imaginary  $s$ , the Chern-Simons path integral, like the Euclidean quantum gravity path integral in any dimension, does not appear to be convergent. Our analysis in this paper will explain how to make sense of this sort of path integral, at least in the Chern-Simons case.

Another motivation is to understand aspects of the Jones polynomial of knots<sup>2</sup> and its generalizations that have not yet been explained from the point of view of three-dimensional gauge theory. The Jones polynomial of a knot was originally defined as a Laurent polynomial in a complex variable  $q$  – a rational function of  $q$  with poles only at  $q = 0$  and  $q = \infty$ . Chern-Simons gauge theory with gauge group  $SU(2)$  (or another compact Lie group) reproduces the Jones polynomial, and its generalizations, at values of  $q$  of the form  $\exp(2\pi i/(k + h))$ , where  $k$  is a positive integer, and  $h$  (which is 2 for  $SU(2)$ ) is the dual Coxeter number of the gauge group. The Jones polynomial and related invariants are determined by their values for these special values of  $q$ , but it would be desirable to have a gauge theory explanation of the existence of a natural analytic continuation. It has generally been expected that

---

<sup>1</sup>And the singularity at  $s = \pm i\ell$  that we will discuss in section 2.4 is at least a cousin of the “chiral point” of three-dimensional gravity, which has been much studied recently [1].

<sup>2</sup>The usual mathematical terminology is that a knot is an embedded circle in a three-manifold (usually a three-sphere) while a union of disjoint embedded circles is called a link. Except in section 4.2.7, all statements about knots in this paper have rather immediate analogs for links, but for brevity we will refer to knots.

the explanation has something to do with the relation between Chern-Simons theories with gauge groups  $SU(2)$  or  $SL(2, \mathbb{C})$ , but a really clear picture has not been available. A better understanding appears to require analytic continuation of Chern-Simons theory beyond the values of the parameters at which the path integral can be most simply understood.

Finally, in the last decade or more, striking new results have been obtained [2–21] about the analytic continuation of invariants related to the Jones polynomial. (A useful review is [17]; partial physics-based explanations with some relevance to the present paper are [6, 18, 19]. While it is impractical to give complete references on Chern-Simons theory, some contributions comparing the asymptotic or in some cases exact behavior of the Chern-Simons partition function to geometry, but not focusing on analytic continuation, are [22–34].) The colored Jones polynomial  $J_n(q)$  of a knot is interpreted in Chern-Simons gauge theory in terms of the expectation value of a Wilson loop operator associated to the  $n$ -dimensional representation of  $SU(2)$ . Let  $q^n = \exp(2\pi i \gamma)$ , and consider the behavior of  $J_n(q)$  in a scaling limit with  $n$  tending to infinity for fixed  $\gamma$ . The behavior of this limit for certain values of  $\gamma$  can be naturally understood in  $SU(2)$  Chern-Simons theory, but many of the more recent results certainly appear to be more directly related to the theory with gauge group  $SL(2, \mathbb{C})$ . For example, for  $\gamma$  near 1, and at least for some knots, the large  $n$  behavior appears to be dominated by an  $SL(2, \mathbb{C})$ -valued flat connection that does not take values in  $SU(2)$ . Although there has not been a completely clear picture of how and why the scaling limit of  $J_n(q)$  should be related to  $SL(2, \mathbb{C})$  Chern-Simons theory, the idea that there is such a relationship has been a fruitful source of new ideas and generalizations [6]. Again, it has been clear that a fuller picture will require a better understanding of the analytic continuation of Chern-Simons theory.

In this paper, we will definitely not obtain complete answers to all of the above questions, and among other things we do not have a general answer concerning the volume conjecture. (It is true, as will become clear, if a certain integer-valued coefficient is always nonzero, but we do not know why this would always be so.) But we do hope to make some things clearer. We show that critical points of the analytically continued Chern-Simons functional should be used to determine – by steepest descent – suitable integration cycles for the quantum path integral. For a given value of the coupling parameters, not all critical points should be included; doing so would lead to immediate contradiction with known facts about the knot polynomials. Instead, one should start with a real integration cycle for real coupling and then, as one varies the coupling into the complex plane, one varies the integration cycle so that the integral remains convergent. In the process, one encounters subtle Stokes phenomena that govern the appearance and disappearance of critical points. The Stokes phenomena can be described, in principle, by solving steepest descent equations that in the context of Chern-Simons theory amount to four-dimensional elliptic differential equations that generalize the instanton equation. This gives a suitable framework for analytically continuing the  $SL(2, \mathbb{C})$  Chern-Simons theory as a function of its coupling parameter, and for better understanding its relation to the  $SU(2)$  theory.

Section 2 of the paper is devoted to a more complete overview of some of these matters.

In section 3, we describe the analytic continuation of oscillatory integrals in finite dimensions and their analysis by steepest descent. This is an application of the framework of [35–40], which is reviewed here for convenience. Various aspects of this framework have been applied to quantum field theory from different points of view [41–45]. Floer theory (for reviews see for instance [46, 47]) is a prototype for the application of Morse theory to middle-dimensional cohomology of function spaces, which appears in our approach to Chern-Simons gauge theory.

In section 4, we apply what we have learned to the analytic continuation of Chern-Simons gauge theory on a generic three-manifold. A striking result emerges: the gradient flow equations relevant to analytic continuation of three-dimensional Chern-Simons theory actually have four-dimensional symmetry.

In section 5, we consider the special case of knot invariants in  $S^3$ , where stronger conclusions are possible because, in the absence of the knot, there are no non-trivial flat connections. The main examples considered in section 5 are the trefoil knot and the figure-eight knot. In each case, we consider the semi-classical limit of the colored Jones polynomial, keeping fixed the conjugacy class of the monodromy of a flat connection around the knot. (In other words, we consider the limit mentioned above of  $n \rightarrow \infty$  with fixed  $\gamma$ .) The technique that we use to describe the Stokes phenomena relies on analyzing the singularities of the moduli space of flat connections – that is, the special values of the monodromy at which flat connections that are generically distinct become coincident. The semiclassical behavior near such a special value is interesting in itself; it can be modeled by a finite-dimensional integral associated to the relevant singularity. We analyze several examples of such singularities and compare the results to what has been found in the mathematical literature in studies of the semiclassical limit of the colored Jones polynomial. For the trefoil knot, though many interesting questions remain, the Stokes phenomena that we analyze suffice to determine the integration cycle of the analytically continued Chern-Simons path integral for all values of  $\gamma$ . For the figure-eight knot, we primarily determine what happens close to the real  $\gamma$  axis. In both cases, our analysis of the Stokes phenomena suffices to explain the validity of the volume conjecture.

Finally, in section 6, we make a few remarks on a four or five-dimensional interpretation of what we have learned, with the possible goal of a new understanding of Khovanov homology.

## 2 Overview

We begin by considering Chern-Simons gauge theory with the gauge group  $G$  a complex Lie group such as  $SL(2, \mathbb{C})$ . Let  $\mathcal{A}$  be a connection on a  $G$ -bundle  $E$  over a three-manifold  $M$ . Such a connection has a complex-valued Chern-Simons invariant

$$W(\mathcal{A}) = \frac{1}{4\pi} \int_M \text{Tr} \left( \mathcal{A} \wedge d\mathcal{A} + \frac{2}{3} \mathcal{A} \wedge \mathcal{A} \wedge \mathcal{A} \right), \quad (2.1)$$

which we have normalized to be gauge-invariant modulo  $2\pi$ . (For  $G = SL(N, \mathbb{C})$ , the symbol  $\text{Tr}$  denotes the trace in the  $N$ -dimensional representation. In general, it denotes a suitably normalized quadratic form on the Lie algebra  $\mathfrak{g}$  of  $G$ .) Just as in the case of a compact group,  $W(\mathcal{A})$  is gauge-invariant modulo  $2\pi$ . The indeterminacy in  $W(\mathcal{A})$  is real, and is the same as it would be if  $G$  were replaced by a maximal compact subgroup  $H$ , because  $G$  is contractible onto  $H$ .<sup>3</sup>

Thus the imaginary part of  $W(\mathcal{A})$  is actually a well-defined real number, while its real part takes values in  $\mathbb{R}/2\pi\mathbb{Z}$ . To obtain a quantum field theory, we wish to start with a classical action  $I$  that is a linear combination of  $\text{Re } W$  and  $\text{Im } W$ . The quantum theory is based upon integrating the expression  $\exp(iI)$ , and for this function to be well-defined,  $I$  must be defined mod  $2\pi$ . Because of the indeterminacy in  $\text{Re } W$ , its coefficient must be an integer, while (if the only desired condition is to ensure that  $\exp(iI)$  is a well-defined complex-valued function) the coefficient of  $\text{Im } W$  may be an arbitrary complex number. The action therefore has the general form

$$I = -s \text{Im } W + \ell \text{Re } W, \quad s \in \mathbb{C}, \ell \in \mathbb{Z}. \quad (2.2)$$

Alternatively, we can write

$$I = \frac{tW}{2} + \frac{\tilde{t}\overline{W}}{2} = \frac{t}{8\pi} \int_M \text{Tr} \left( \mathcal{A} \wedge d\mathcal{A} + \frac{2}{3} \mathcal{A} \wedge \mathcal{A} \wedge \mathcal{A} \right) + \frac{\tilde{t}}{8\pi} \int_M \text{Tr} \left( \overline{\mathcal{A}} \wedge d\overline{\mathcal{A}} + \frac{2}{3} \overline{\mathcal{A}} \wedge \overline{\mathcal{A}} \wedge \overline{\mathcal{A}} \right), \quad (2.3)$$

with

$$t = \ell + is, \quad \tilde{t} = \ell - is. \quad (2.4)$$

There is more to life, however, than making sure that  $\exp(iI)$  is a well-defined complex-valued function. We also want to integrate this function over the space  $\mathcal{Y}$  of gauge fields

---

<sup>3</sup>The indeterminacy in  $W(\mathcal{A})$  comes from its change in going around a loop in the space of connections modulo gauge transformations, and this change only depends on the homotopy class of the loop. Since  $G$  is contractible to  $H$ , any family of  $\mathfrak{g}$ -valued connections on  $M$  can be deformed to a family of  $\mathfrak{h}$ -valued connections, where  $\mathfrak{g}$ ,  $\mathfrak{h}$  are the Lie algebras of  $G$  and  $H$ . So the indeterminacy in the Chern-Simons function for gauge group  $G$  reduces to that for gauge group  $H$ .



modulo gauge transformations. This makes sense straightforwardly only if  $s$  is real, for then  $I$  is real and the integrand  $\exp(iI)$  of the path integral is bounded. The path integral

$$\int_{\mathcal{Y}} D\mathcal{A} D\bar{\mathcal{A}} \exp(iI) \quad (2.5)$$

is then an infinite-dimensional oscillatory integral. In this respect, it is like any Lorentz signature Feynman integral (though here we are doing topological field theory rather than quantum field theory on a manifold of Lorentz signature). If  $s$  is not real, the path integral is not convergent; our task is to make sense of it anyway.

## 2.1 Finite-Dimensional Analogs

A much simpler prototype of this problem is a finite-dimensional oscillatory integral such as the one that defines the Airy function:

$$Z_\lambda = \int_{-\infty}^{\infty} dx \exp(i\lambda(x^3/3 - x)). \quad (2.6)$$

For real  $\lambda$ , this integral converges, though not absolutely so, because of rapid oscillations at infinity; when  $\lambda$  becomes complex, it must be defined by a process of analytic continuation.

The Airy integral will serve as a useful practice case in section 3. The following slightly more elaborate example is also useful. Introduce a complex variable  $z$  and a complex-valued polynomial

$$g(z) = \sum_{j=0}^n a_j z^j. \quad (2.7)$$

Now consider the integral

$$Z_g = \int |d^2 z| \exp\left(g(z) - \overline{g(z)}\right). \quad (2.8)$$

This again is a convergent oscillatory integral and a closer analog of complex Chern-Simons theory, with  $-ig(z)$  and  $i\overline{g(z)}$  corresponding to the terms  $tW$  and  $\tilde{t}\overline{W}$  in the action (2.3).

We stress that no contour integral is intended in (2.8) – a contour integral could scarcely be intended here as the integrand is not holomorphic! Rather, if  $z = u + iv$ , with real  $u, v$ , we integrate separately over  $u$  and  $v$ , the integration measure being  $|d^2 z| = 2 du dv$ .

A finite-dimensional analog of analytically continuing Chern-Simons theory with complex gauge group as a function of  $s$  is to analytically continue the Airy function  $Z_\lambda$  to complex

$\lambda$ . An even closer analog is to analytically continue the integral  $Z_g$  to let  $\bar{g}$  be independent of  $g$ . By this, we mean the following. Define a new polynomial

$$\tilde{g}(z) = \sum_{j=0}^n \tilde{a}_j z^j, \quad (2.9)$$

and generalize the integral  $Z_g$  to

$$Z_{g,\tilde{g}} = \int |d^2 z| \exp(g(z) - \tilde{g}(\bar{z})). \quad (2.10)$$

Then  $Z_{g,\tilde{g}}$  coincides with the original  $Z_g$  if  $\tilde{g} = \bar{g}$  (by which we mean that  $\tilde{a}_j = \bar{a}_j$  for all  $j$ ).

The natural way to analytically continue the Airy function in  $\lambda$  is to let  $x$  become a complex variable and to deform the integration contour away from the real axis so that the integral remains convergent as  $\lambda$  varies. Similarly, to extend the definition of  $Z_{g,\tilde{g}}$  to  $\tilde{g} \neq \bar{g}$ , we analytically continue, treating  $z$  and  $\bar{z}$  – or equivalently  $u$  and  $v$ , defined by  $z = u + iv$ ,  $\bar{z} = u - iv$  – as independent complex variables. Denoting the independent complex variables as  $z$  and  $\tilde{z}$  (rather than  $z$  and  $\bar{z}$ ), the analytically continued integral is

$$Z_{g,\tilde{g}} = \int_{\mathcal{C}} dz d\tilde{z} \exp(g(z) - \tilde{g}(\tilde{z})). \quad (2.11)$$

The integral is over a two-dimensional real integration cycle  $\mathcal{C}$ , which is the real slice  $\tilde{z} = \bar{z}$  if  $\tilde{g} = \bar{g}$ , and in general must be deformed as  $g$  and  $\tilde{g}$  vary so that the integral remains convergent.

This then is the framework that we will have to imitate to understand the analytic continuation of Chern-Simons theory with a complex gauge group. Just as we promoted  $\bar{z}$  to a complex variable  $\tilde{z}$  that is independent of  $z$ , we will have to promote  $\bar{\mathcal{A}}$  to a new  $G$ -valued connection  $\tilde{\mathcal{A}}$  that is independent of  $\mathcal{A}$ . Thus we consider the classical theory with independent  $G$ -valued connections  $\mathcal{A}$  and  $\tilde{\mathcal{A}}$  (we recall that  $G$  is a complex Lie group such as  $SL(2, \mathbb{C})$ ) and action

$$I(\mathcal{A}, \tilde{\mathcal{A}}) = \frac{t}{8\pi} \int_M \text{Tr} \left( \mathcal{A} \wedge d\mathcal{A} + \frac{2}{3} \mathcal{A} \wedge \mathcal{A} \wedge \mathcal{A} \right) + \frac{\tilde{t}}{8\pi} \int_M \text{Tr} \left( \tilde{\mathcal{A}} \wedge d\tilde{\mathcal{A}} + \frac{2}{3} \tilde{\mathcal{A}} \wedge \tilde{\mathcal{A}} \wedge \tilde{\mathcal{A}} \right). \quad (2.12)$$

Then we have to find an appropriate integration cycle  $\mathcal{C}$  in the path integral

$$\int_{\mathcal{C}} D\mathcal{A} D\tilde{\mathcal{A}} \exp(iI(\mathcal{A}, \tilde{\mathcal{A}})). \quad (2.13)$$

The cycle  $\mathcal{C}$  must be equivalent to  $\tilde{\mathcal{A}} = \bar{\mathcal{A}}$  if  $s$  is real, and in general must be obtained by deforming that one as  $s$  varies so that the integral remains convergent. (We also have to learn how to deal with gauge invariance when  $\mathcal{A}$  and  $\tilde{\mathcal{A}}$  are treated as independent.)

## 2.2 Role Of Morse Theory

Finding a suitable integration cycle in the infinite-dimensional path integral of Chern-Simons gauge theory may seem hopeless at first sight. What saves the day is that suitable integration contours can be found by Morse theory and steepest descent (or in a different language, via Picard-Lefschetz theory). The facts we need for understanding oscillatory integrals have been described in [35–40] (and the aspiration of applying this approach to Feynman integrals is stated in [37], p. 321). Section 3 of the present paper is devoted to a hopefully self-contained exposition of this approach, with an emphasis on the aspects that we need in Chern-Simons theory. The basic idea of the theory is to associate potential integration cycles to components of the critical point set of the extended action  $I(\mathcal{A}, \tilde{\mathcal{A}})$ .

What makes it practical to use this formalism in Chern-Simons theory is that the critical points are accessible. As long as  $t$  and  $\tilde{t}$  are both non-zero, the critical point equation derived from the analytically continued action (2.12) is

$$\mathcal{F} = \tilde{\mathcal{F}} = 0, \quad (2.14)$$

where  $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$  and  $\tilde{\mathcal{F}} = d\tilde{\mathcal{A}} + \tilde{\mathcal{A}} \wedge \tilde{\mathcal{A}}$  are the curvatures of the two connections. Hence a critical point is just a pair of flat connections, corresponding to a pair of homomorphisms from the fundamental group of  $M$  to the complex Lie group  $G$ . (We postpone to section 4 an important detail about the value of the Chern-Simons invariant at a critical point.) Hence in Chern-Simons theory on a given three-manifold  $M$ , the critical points can be effectively described; this is typically not true for oscillatory Feynman integrals.

For simplicity, we assume until section 3.3.2 that the critical points are isolated points. Let  $\Sigma$  be the set of critical points and denote the critical points as  $p_\sigma$ ,  $\sigma \in \Sigma$ . Then to each  $p_\sigma$ , we attach an integration cycle  $\mathcal{J}_\sigma$  that is obtained by downward or gradient flow from  $\sigma$ , taking the real part of the action as a Morse function. (Details are explained in section 3.) The  $\mathcal{J}_\sigma$ , which are known as Lefschetz thimbles, are constructed so that the integral over  $\mathcal{J}_\sigma$  always converges, regardless of the choice of  $\sigma$  or of the parameters in the action. Any suitable integration cycle  $\mathcal{C}$  for the sort of integral that we are interested in is equivalent to a linear combination of the  $\mathcal{J}_\sigma$  with integer coefficients:

$$\mathcal{C} = \sum_{\sigma} \mathbf{n}_\sigma \mathcal{J}_\sigma, \quad \mathbf{n}_\sigma \in \mathbb{Z}. \quad (2.15)$$

The coefficients  $\mathbf{n}_\sigma$  can themselves be computed via Morse theory. To do this, one attaches to each critical point  $p_\sigma$  another cycle  $\mathcal{K}_\sigma$  by upwards gradient flow starting from  $p_\sigma$ . The expansion coefficients in eqn. (2.15) can be computed as the intersection pairing of the integration cycle  $\mathcal{C}$  with the  $\mathcal{K}_\sigma$ :

$$\mathbf{n}_\sigma = \langle \mathcal{C}, \mathcal{K}_\sigma \rangle. \quad (2.16)$$

One might naively think that all critical points would contribute to the integral with  $\mathbf{n}_\sigma = 1$  (or with  $\mathbf{n}_\sigma = \pm 1$ , since in general the  $\mathcal{J}_\sigma$  do not have natural orientations), but this cannot be so. For example, in Chern-Simons theory with real  $s$ , there are complex critical points at which the action  $I$  has a negative imaginary part; their inclusion would lead to a pathologically wrong exponential growth of the partition function for large real  $s$ . In fact, for real  $s$ , with  $\mathcal{C}$  taken to be the real cycle  $\mathcal{C}_\mathbb{R}$  defined by  $\tilde{\mathcal{A}} = \overline{\mathcal{A}}$ , real critical points (whose contributions to the path integral are oscillatory) all have  $\mathbf{n}_\sigma = 1$ , but critical points that would make exponentially large contributions to the path integral have  $\mathbf{n}_\sigma = 0$ . These statements will be deduced from (2.16). However, there is no simple formula for  $\mathbf{n}_\sigma$  for critical points whose contributions to the real  $s$  path integral are exponentially small.

## 2.3 Stokes Phenomena

Once one expresses the integration cycle  $\mathcal{C}$  in terms of critical point cycles for real values of the Chern-Simons coupling constant  $s$  (or of the appropriate parameters in the other examples) via a formula  $\mathcal{C} = \sum_\sigma \mathbf{n}_\sigma \mathcal{J}_\sigma$ , the next step is to make an analytic continuation in  $s$ . There is no problem in defining the critical points  $p_\sigma$  and the corresponding Lefschetz thimbles  $\mathcal{J}_\sigma$  for any generic value of  $s$ . In fact, in Chern-Simons theory, the critical point equation (2.14) is independent of  $s$  (except at  $s = \pm i\ell$ , where  $t$  or  $\tilde{t}$  vanishes). The  $\mathcal{J}_\sigma$  do depend on  $s$  (their definition by steepest descent makes use of the action, which depends on  $s$ ), but their definition ensures that the path integral over  $\mathcal{J}_\sigma$  is always convergent, even when  $s$  has an imaginary part.

One can find the right coefficients in the expansion  $\mathcal{C} = \sum_\sigma \mathbf{n}_\sigma \mathcal{J}_\sigma$  by starting at real  $s$ , where one knows what to do, and then continuously evolving in  $s$ . Here, however, one runs into the main subtlety of the theory: Stokes phenomena. As one varies  $s$ , one crosses Stokes curves at which both the Lefschetz thimbles  $\mathcal{J}_\sigma$  and the coefficients  $\mathbf{n}_\sigma$  jump, in such a way that the homology cycle  $\sum_\sigma \mathbf{n}_\sigma \mathcal{J}_\sigma$  is locally constant. Such jumping can occur when there are paths of steepest descent that connect two critical points. The flow by steepest descent has a conserved quantity (the imaginary part of the holomorphic function whose real part is taken as the Morse function), and a flow between two critical points is possible only when this quantity takes equal values at the two critical points. This is reviewed in section 3.

After taking the Stokes phenomena into account, one can obtain a recipe for how the  $\mathbf{n}_\sigma$  evolve as  $s$  varies in the complex plane. This then gives, finally, a way to define the analytic continuation in  $s$  of the Chern-Simons path integral for the case that the gauge group is a complex Lie group.

## 2.4 Singularities Of The Analytically Continued Theory

At this stage, it is very illuminating to ask what sort of singularities the analytically continued theory has as a function of  $s$ . Generically, the analytic continuation of an oscillatory integral grows exponentially as the appropriate complex parameter goes to infinity in certain directions in the complex plane. The appropriate parameters are  $s$  for complex Chern-Simons theory,  $\lambda$  for the Airy function, and the coefficients of the polynomials  $g$  and  $\tilde{g}$  in our other example.

Because of this exponential growth, the analytically continued Chern-Simons path integral has an essential singularity at  $s = \infty$ . Generically, there is in addition an ordinary monodromy around  $s = \infty$ ; it results from the fact that, by virtue of the Stokes phenomena, the integration contour  $\mathcal{C}$  does not return to itself under continuation around  $s = \infty$ .

Apart from an essential singularity at infinity, will analytically continued Chern-Simons theory have other singularities as a function of  $s$ ? The answer to this question is that one should expect singularities at  $s = \pm i\ell$ , where the action (2.3) becomes holomorphic or antiholomorphic. An oversimplified prototype of this singularity is the Gaussian integral

$$\int |dz \wedge d\bar{z}| \exp(az^2 + b\bar{z}^2) = \frac{\pi}{\sqrt{-ab}}, \quad (2.17)$$

with singularities at  $a = 0$  or  $b = 0$ . (The integral is defined by analytic continuation starting from an oscillatory Gaussian integral at  $b = -\bar{a}$ .)

This example is oversimplified because the analytically continued exponent, which is  $az^2 + b\bar{z}^2$ , has only a unique critical point, the one at  $z = \tilde{z} = 0$ . In Chern-Simons theory, there are in general many critical points  $p_\sigma$ . In general there are distinct critical points that have the same action at  $s = \pm i\ell$ , and as a result many Stokes curves pass through the points  $s = \pm i\ell$ . Hence the singularities at  $s = \pm i\ell$  will be much more complicated than in the Gaussian example. These singularities will not be analyzed in this paper, but it will not be surprising if the analytic continuation of Chern-Simons theory in general should be defined on the universal cover of the complex  $s$ -plane, with the points at  $s = \pm i\ell$  (and  $\infty$ ) omitted. Thus the partition function would be really a function not of  $s$  but of a certain dilogarithm of  $s$ .

## 2.5 Analytic Continuation For Compact Gauge Groups

Now we will discuss the analytic continuation of Chern-Simons theory with a compact gauge group. We consider it first in a naive way and then in relation to the theory with a complex gauge group.

In Chern-Simons gauge theory on a three-manifold  $M$  with a compact gauge group  $H$ , the gauge field is a connection  $A$  on an  $H$ -bundle  $E \rightarrow M$ . In terms of the Chern-Simons function  $W(A)$ , which we normalize as usual to be well-defined modulo  $2\pi\mathbb{Z}$ , the action is  $I = kW$ , where  $k$  is taken to be an integer so that the integrand  $\exp(iI) = \exp(ikW)$  of the path integral is well-defined. Then one defines the path integral as the integral over the space<sup>4</sup>  $\mathcal{U}$  of connections mod gauge transformations:

$$Z_H(k) = \int_{\mathcal{U}} DA \exp(ikW(A)). \quad (2.18)$$

In the presence of a knot  $K$  labeled by a representation  $R$ , one includes a holonomy factor and considers the integral:

$$Z_H(K; k) = \int_{\mathcal{U}} DA \exp(ikW(A)) \operatorname{Tr}_R P \exp \left( - \oint_K A \right). \quad (2.19)$$

As written, the integral does not make sense unless  $k$  is an integer. Indeed, as  $W(A)$  is defined only modulo  $2\pi$ , the expression  $\exp(ikW(A))$  makes sense as a function on  $\mathcal{U}$  if and only if  $k \in \mathbb{Z}$ . Can we evade this limitation? A naive way to try to do so is to replace  $\mathcal{U}$  by its universal cover<sup>5</sup>  $\widehat{\mathcal{U}}$ .  $\widehat{\mathcal{U}}$  is the quotient of the space of connections by the group of topologically trivial gauge transformations. The expression  $\exp(ikW(A))$  is invariant under such gauge transformations, so it makes sense as a function on  $\widehat{\mathcal{U}}$ . However, its integral vanishes for  $k \notin \mathbb{Z}$ , since the integrand transforms with a non-trivial phase under the deck transformations associated to the cover  $\widehat{\mathcal{U}} \rightarrow \mathcal{U}$ .

To cure this problem, we complexify the gauge field, replacing  $A$  by a complex-valued connection  $\mathcal{A}$ .  $\mathcal{A}$  is now a connection on an  $H_{\mathbb{C}}$ -bundle  $E_{\mathbb{C}} \rightarrow M$ . We replace  $\mathcal{U}$  and  $\widehat{\mathcal{U}}$  by their complexifications  $\mathcal{U}_{\mathbb{C}}$  and  $\widehat{\mathcal{U}}_{\mathbb{C}}$ , which respectively parametrize complex-valued connections modulo complex-valued gauge transformations or modulo topologically trivial complex-valued gauge transformations. Finally, in  $\widehat{\mathcal{U}}_{\mathbb{C}}$ , we pick an integration cycle  $\mathcal{C}$  and define the integral

$$Z_{H,\mathcal{C}}(k) = \int_{\mathcal{C}} D\mathcal{A} \exp(ikW(\mathcal{A})). \quad (2.20)$$

---

<sup>4</sup> A variant in which one takes  $\mathcal{U}$  to be the quotient of the space of connections by the group of based gauge transformations is also useful. (A based gauge transformation is one that equals 1 at a prescribed point; the advantage of dividing only by this group is that it acts freely, so dividing by it produces no singularities.) In this variant, one must divide the path integral by a constant factor, the volume of the residual group of gauge transformations. Another important variant arises in working out the contribution to the path integral from a given gauge orbit of flat connections on  $M$ ; here one may take  $\mathcal{U}$  to be a transversal to the given orbit in the space of all connections on  $E$ . When we speak loosely of  $\mathcal{U}$  as the space of connections mod gauge transformations, we refer to any of these variants that may be most convenient in a given context.

<sup>5</sup>To be more precise, we can take  $\widehat{\mathcal{U}}$  to be the smallest cover of  $\mathcal{U}$  on which the Chern-Simons function  $W(A)$  makes sense as a real-valued function. If  $H$  is connected and simply-connected, this is the same as the universal cover.

(When a knot  $K$  is present, we denote the corresponding integral as  $Z_{H,\mathcal{C}}(K; k)$ .) The analytic continuation of the path integral measure  $DA$  on  $\widehat{\mathcal{U}}$  is a holomorphic volume form on  $\widehat{\mathcal{U}}_{\mathbb{C}}$  that we continue to denote simply as  $DA$ . In section 2.7.1, it will be expressed in terms of the holomorphic volume form  $D\mathcal{A}$  used in the complex path integral (2.13).

We will pick  $\mathcal{C}$  so that the integral (2.20) converges even when  $k$  becomes a complex number. It will be constructed from Lefschetz thimbles in the usual way,

$$\mathcal{C} = \sum_{\sigma} \mathbf{n}_{\sigma} \mathcal{J}_{\sigma}. \quad (2.21)$$

Here the  $p_{\sigma}$  correspond to flat connections  $\mathcal{A}_{\sigma}$  with structure group  $H_{\mathbb{C}}$ , each with a chosen lift to  $\widehat{\mathcal{U}}_{\mathbb{C}}$ . (The sum in (2.21) will generally be a finite sum, but see the end of section 5.2.4 for an exception.) The analytically continued path integral will therefore be

$$Z_H(k) = \sum_{\sigma} \mathbf{n}_{\sigma} Z_{H,\sigma}(k) \quad (2.22)$$

with

$$Z_{H,\sigma}(k) = \int_{\mathcal{J}_{\sigma}} DA \exp(ikW(\mathcal{A})). \quad (2.23)$$

A finite sum (2.21) is not invariant under the deck transformations of the cover  $\widehat{\mathcal{U}}_{\mathbb{C}} \rightarrow \widehat{\mathcal{U}}$ , so it is not possible to use those deck transformations to show that  $Z_{H,\mathcal{C}}(K, k) = 0$  for  $k \notin \mathbb{Z}$ . (Invariance under deck transformations would require including every lift of a given  $\mathcal{A}_{\sigma}$ .) As a cycle in the appropriate homology,  $\mathcal{C}$  will be locally constant in  $k$  (Stokes jumping of cycles  $\mathcal{J}_{\sigma}$  is compensated by jumping of coefficients  $\mathbf{n}_{\sigma}$ ), ensuring that  $Z_{H,\mathcal{C}}(K, k)$  will be holomorphic in  $k$ .

### 2.5.1 Downward Flow Again

To get an analytic continuation of the original Chern-Simons integral (2.18),  $\mathcal{C}$  should obey one more condition: it should be chosen so that  $Z_{H,\mathcal{C}}(k) = Z_H(k)$  for integer  $k$ .

In a finite dimensional problem of this type, a suitable  $\mathcal{C}$  can be found as follows. (For a prototype, see the review in section 3.5 of the analytic continuation of the Bessel function.) First, lift the real cycle  $\mathcal{U} \subset \mathcal{U}_{\mathbb{C}}$  to a non-closed chain  $\mathcal{C}_0 \subset \widehat{\mathcal{U}}_{\mathbb{C}}$  as follows. Pick a real number  $\alpha$  and define a point in  $\mathcal{C}_0$  to be a real connection  $A$  whose Chern-Simons function takes values in the closed interval  $[\alpha, \alpha + 2\pi]$ . (We are using the fact that on  $\widehat{\mathcal{U}}_{\mathbb{C}}$ , the Chern-Simons function is an ordinary complex-valued function.) We have to use a closed interval here or  $\mathcal{C}_0$  would not be closed. Though closed,  $\mathcal{C}_0$  is not a cycle because it has a non-zero boundary  $\partial\mathcal{C}_0$

– consisting of real connections with Chern-Simons function  $\alpha$  or  $\alpha + 2\pi$ . Of course,  $\partial\mathcal{C}_0$  is a union of two disjoint pieces, which we will loosely call components (we do not need to know if they are connected), corresponding to those two values of the Chern-Simons function.

We want to complete  $\mathcal{C}_0$  to a candidate integration cycle  $\mathcal{C} \subset \widehat{\mathcal{U}}_{\mathcal{C}}$ . For this, we add to  $\mathcal{C}_0$  an additional piece  $\mathcal{C}_1$  defined as follows. We define  $\mathcal{C}_1$  by downward Morse flow from the boundary  $\partial\mathcal{C}_0$ , with respect to the usual Morse function  $h = \text{Re}(ikW(\mathcal{A}))$ . Assume first that downward flow from  $\partial\mathcal{C}_0$  leads to no critical points. Then  $\mathcal{C}_1$  is closed and, if we orient correctly its two components, then  $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1$  has no boundary and is the integration cycle we want. (In the prototype of fig. 7 in section 3.5, the two components of  $\mathcal{C}_1$  correspond to the two half-lines that start at  $w = 0$  and  $w = 2\pi i$ .)

$\mathcal{C}_1$  fails to be closed only if a downward flow from  $\partial\mathcal{C}_0$  approaches a critical point for  $t \rightarrow +\infty$ . (This critical point is then in the closure of  $\mathcal{C}_1$ , but not in  $\mathcal{C}_1$  itself.) This possibility is severely limited because the downward flow equations have a conserved quantity, which for Chern-Simons theory is  $\text{Im}(ikW(\mathcal{A}))$ , while in addition,  $\text{Re}(ikW(\mathcal{A}))$  decreases along a flow. So  $\mathcal{C}_1$  may fail to be closed only if there is a critical point at which  $W(\mathcal{A})$  has positive imaginary part and real part  $\alpha$  or  $\alpha + 2\pi$ . For a generic choice of  $\alpha$ , there is no such critical point and the above definition works and gives a suitable cycle  $\mathcal{C}$ .

Now consider the integral

$$\int_{\mathcal{C}} DA \exp(ikW(\mathcal{A})) = \int_{\mathcal{C}_0} DA \exp(ikW(\mathcal{A})) + \int_{\mathcal{C}_1} DA \exp(ikW(\mathcal{A})). \quad (2.24)$$

Here  $\mathcal{C}_1$  consists of two components (downward flow from real connections with  $W(\mathcal{A})$  equal to  $\alpha$  or  $\alpha + 2\pi$ ) that differ from each other by a deck transformation that shifts  $W(\mathcal{A})$  by  $2\pi$ . For  $k \in \mathbb{Z}$ , the integrand  $\exp(ikW(\mathcal{A}))$  is invariant under this deck transformation, but on the other hand, the two components of  $\mathcal{C}_1$  are oppositely oriented (that is, the deck transformation maps the orientation of one to minus the orientation of the other). So for  $k \in \mathbb{Z}$ , we have  $\int_{\mathcal{C}_1} DA \exp(ikW(\mathcal{A})) = 0$  and  $\int_{\mathcal{C}} DA \exp(ikW(\mathcal{A})) = \int_{\mathcal{C}_0} DA \exp(ikW(\mathcal{A}))$ . But for  $k \in \mathbb{Z}$ , the function  $\exp(ikW(\mathcal{A}))$  is a pullback from the original space  $\mathcal{U}_{\mathcal{C}}$  of complex-valued connections modulo gauge transformations, and the cycle  $\mathcal{C}_0$  pushes down in  $\mathcal{U}_{\mathcal{C}}$  to the original real integration cycle  $\mathcal{U}$ . In short,  $\int_{\mathcal{C}} DA \exp(ikW(\mathcal{A}))$  is equivalent to the original Chern-Simons path integral when  $k$  is an integer. We have accomplished our goal of expressing the Chern-Simons path integral  $Z_H(k)$  with a compact gauge group in a form that makes sense when  $k$  is not an integer.

$\mathcal{C}$  is an unfamiliar sort of integration cycle for a path integral, but actually, it is straightforward in concept to use (2.16) to re-express  $\mathcal{C}$  in terms of Lefschetz thimbles. We just count the upward flows from a given critical point  $\mathcal{A}_{\sigma}$  to  $\mathcal{C}$ . If  $k$  is real, the flow equations conserve  $\text{Im}(iW(\mathcal{A}))$ , so, given the condition that we have imposed on  $\alpha$ ,  $\mathbf{n}_{\sigma}$  is nonzero only for critical points for which  $\text{Im}(iW(\mathcal{A}))$  takes values in the open interval  $(\alpha, \alpha + 2\pi)$ .



### 2.5.2 Indeterminacy of $\mathcal{C}$

The only problem with the analytic continuation that we have obtained for the Chern-Simons path integral is that it is not unique.

The choice of  $\mathcal{C}$  and therefore the analytic continuation of the Chern-Simons path integral  $Z_H(k)$  will in general jump in crossing the exceptional values of  $\alpha$  at which  $\mathcal{C}_1$  fails to be closed. This is related to a very basic fact. The contribution of any given critical point  $p_\sigma$  to the path integral can be multiplied by  $\exp(2\pi imk)$ , for any integer  $m$ , without changing its value for  $k \in \mathbb{Z}$ . This multiplicative factor has a simple interpretation: it results from shifting the Chern-Simons invariant of the flat connection  $\mathcal{A}_\sigma$  by  $2\pi m$ . So, without more input, we cannot expect analytic continuation in  $k$  to be unique.

In a situation (such as that of section 5) in which the Chern-Simons path integral depends on another parameter  $\gamma$  in addition to  $k$ , it may not be possible to make our almost canonical choice of  $\mathcal{C}$  uniformly in  $\gamma$ . The reason is that it may not be possible to choose  $\alpha$  so that it varies continuously with  $\gamma$  and never coincides with the value of  $\text{Im}(iW(\mathcal{A}))$  for any flat connection  $\mathcal{A}$ . If we choose  $\mathcal{C}$  as above for some range of  $\gamma$ , then upon varying  $\gamma$  and letting  $\mathcal{C}$  evolve continuously, we will arrive at a more general  $\mathcal{C}$  and hence a more general form of the analytic continuation.

A rather general  $\mathcal{C}$  would entail multiplying the contribution to the path integral from each representation of the fundamental group by a finite sum

$$\sum_{m \in \mathbb{Z}} a_m \exp(2\pi imk), \quad (2.25)$$

where the integers  $a_m$  are almost all zero and  $\sum_m a_m = 1$ . In general, the  $a_m$  may be different for different representations of the fundamental group. A function such as this will naturally arise in an example in section 5.2.3.

### 2.5.3 Knots

There is an important case in which there is a completely natural choice of the analytic continuation. This is the case of knots or links in  $S^3$ . As is known from any of their standard characterizations, the usual Wilson loop invariants for knots or links can be analytically continued to a Laurent polynomial in  $q = \exp(2\pi i/(k+h))$  and with this analytic continuation, they have no essential singularity at  $k = \infty$ . Since the function  $\exp(2\pi ik)$  does have an essential singularity at  $k = \infty$ , the cycle  $\mathcal{C}$  used in analytically continuing the knot invariants is unique if one wants the analytically continued function to be free of an essential singularity at infinity.

Perhaps we should spell out a detail concerning the last paragraph. The Jones polynomial and its generalizations are usually normalized to equal 1 for the unknot. From the point of view of quantum field theory, this definition amounts to defining the Jones polynomial of a given knot  $K$  by analytic continuation of a ratio  $Z_H(K; k)/Z_H(K_0; k)$ , where  $Z_H(K; k)$  was defined in (2.19) and  $K_0$  is the unknot. The ratio is certainly 1 if  $K = K_0$ . It is such ratios that analytically continue to Laurent polynomials in  $q$ . The numerator and denominator, when taken separately, have analytic continuations in  $k$  with no essential singularity at  $k = \infty$ , but they are not Laurent polynomials in  $q$ . For example, for  $H = SU(2)$ , the denominator is  $Z_H(K_0; k) = \sqrt{2/(k+2)} \sin(\pi/(k+2))$ , which is not a Laurent polynomial in  $q$ , but has no essential singularity at  $k = \infty$ . The theory developed in the present paper applies more naturally to a path integral than to a ratio of path integrals, and therefore we will simply study the functions  $Z_H(K; k)$  rather than ratios of such functions.

## 2.6 Comparing Compact And Complex Lie Groups

If we do not require that the generalized integral  $Z_{H,\mathcal{C}}(k)$  should agree with the original Chern-Simons integral  $Z_H(k)$  when  $k$  is an integer, we have much more freedom in the choice of the cycle  $\mathcal{C}$ . We can take  $\mathcal{C}$  to be any integer linear combination of Lefschetz thimbles

$$\mathcal{C} = \sum_{\sigma \in \Sigma} \mathbf{n}_\sigma \mathcal{J}_\sigma, \quad (2.26)$$

with coefficients  $\mathbf{n}_\sigma$  that jump in crossing Stokes lines to ensure that the homology cycle  $\mathcal{C}$  varies smoothly.

The functions  $Z_{H,\mathcal{C}}(k)$  of eqn. (2.20), as  $\mathcal{C}$  runs over the possible choices, form a vector space  $\mathcal{V}$ . (This space is actually infinite-dimensional when one allows for the choices of lifts to  $\mathbb{C}$  of Chern-Simons invariants of critical points.) What are these functions good for, apart from analytic continuation of Chern-Simons theory with gauge group  $H$ ? They play a role somewhat analogous to conformal blocks in two-dimensional conformal field theory. Let us compare Chern-Simons theory of the compact group  $H$  to Chern-Simons theory of the complex Lie group  $G = H_{\mathbb{C}}$ . For reasons that will become clear, we consider first the case of a three-manifold  $M$  without knots.

As we have explained, the analytic continuation of the complex Chern-Simons theory as a function of its coupling parameters is made using the same machinery as for a compact Lie group. However, the complexification of a complex Lie group  $G$  is isomorphic to the product  $G \times G$ , and accordingly, the analytic continuation of Chern-Simons theory of gauge group  $G$  is described by a pair of complex connections  $\mathcal{A}, \tilde{\mathcal{A}}$ . This assertion was part of the derivation of eqn. (2.12), where we also showed that the analytically continued action is the sum of a Chern-Simons action for  $\mathcal{A}$  and a Chern-Simons action for  $\tilde{\mathcal{A}}$ . We can thus build

an integration cycle for the analytically continued  $G$ -valued theory by taking the product of two integration cycles for the analytically continued  $H$ -valued theory. Concretely, we can consider an integration cycle for the analytically continued  $G$ -valued theory of the form  $\mathcal{J}_\sigma \times \mathcal{J}_\tau$ , where  $\mathcal{J}_\sigma$  and  $\mathcal{J}_\tau$  are Lefschetz thimbles for the analytically continued  $H$ -valued theory. This gives a basis for the space of appropriate cycles, and hence the partition function  $Z_G(s, \ell)$  of the  $G$ -valued theory can be expanded as a bilinear form in the functions  $Z_{H,\sigma}(k)$  of eqn. (2.23):

$$Z_G(s, \ell) = \sum_{\sigma, \tau} \mathbf{n}_{\sigma, \tau} Z_{H,\sigma}(k_1) Z_{H,\tau}(k_2). \quad (2.27)$$

The coefficients  $\mathbf{n}_{\sigma, \tau}$  jump in the usual way in crossing Stokes curves.

From the formulas (2.3) and (2.4) for the analytically continued action, one might expect the coupling parameters to be  $k_1 = t/2 = (\ell + is)/2$ ,  $k_2 = \tilde{t}/2 = (\ell - is)/2$ . Actually there is a shift by the dual Coxeter number  $h$  of  $H$  and one has

$$k_1 = t/2 - h, \quad k_2 = \tilde{t}/2 - h. \quad (2.28)$$

This shift enters in comparing the integration measures  $D\mathcal{A}$  and  $DA$  used in defining  $Z_G$  and  $Z_H$ , or, differently put, it reflects the fact that in Chern-Simons theory with a compact gauge group  $H$ , a one-loop correction shifts the effective value of  $k$  in many formulas to  $k + h$ , while as shown in [48], there is no such shift for a complex Lie group. For more on this, see section 2.7.1.

### 2.6.1 Knots Again

It is instructive to consider what happens to this formalism if we specialize to knots in  $S^3$ , again in the theory with gauge group  $G = H_\mathbb{C}$ . In the presence of a knot  $K$ , we must include in the integrand of the path integral an extra factor that is the holonomy of the connection  $\mathcal{A}$  around  $K$ , in some representation  $R$ . What happens depends very much on what sort of representation we pick. For example, one might pick an irreducible unitary representation of  $G$ ; these of course are infinite-dimensional. In this case, the simplification that will be explained momentarily does not arise.

We will consider here a finite-dimensional and irreducible but non-unitary representation that is related to an irreducible representation of the compact group  $H$ . The action of  $H$  on an irreducible representation  $R$  can be extended to an action of  $G = H_\mathbb{C}$  in two ways; in one case the matrix elements of the representations are holomorphic functions on  $G$  and in the other case they are antiholomorphic. We will consider the case of a holomorphic representation. Given a knot  $K$  and a  $G$ -valued connection  $\mathcal{A}$ , holomorphy means after analytic continuation that the trace of the holonomy around  $K$  is a function only of  $\mathcal{A}$

and not of  $\tilde{\mathcal{A}}$ ; as usual we denote this trace as  $\text{Tr}_R P \exp(-\oint_K \mathcal{A})$ . Now consider the path integral:

$$Z_G(K; t, \tilde{t}) = \int_{\mathcal{C}} D\mathcal{A} D\tilde{\mathcal{A}} \exp\left(itW(\mathcal{A})/2 + i\tilde{t}W(\tilde{\mathcal{A}})/2\right) \text{Tr}_R P \exp\left(-\oint_K \mathcal{A}\right). \quad (2.29)$$

We want to express this integral in terms of Lefschetz thimbles. So consider the equations for a critical point. (We will here proceed somewhat informally; the details of how to include the holonomy factor in the Morse theory framework are explained in section 4.2.) Since the holonomy factor only depends on  $\mathcal{A}$ , the equation for  $\tilde{\mathcal{A}}$  is unaffected by the presence of the knot. This equation says simply that the curvature of  $\tilde{\mathcal{A}}$  vanishes, and as  $S^3$  is simply-connected, this means that up to a gauge transformation we can take  $\tilde{\mathcal{A}} = 0$ .

Consequently, there is essentially only one possible integration cycle for  $\tilde{\mathcal{A}}$ ; it is the Lefschetz thimble  $\tilde{\mathcal{Q}}$  associated to the critical point at  $\tilde{\mathcal{A}} = 0$ . Hence, the integration cycle  $\mathcal{C}$  is simply the product of  $\tilde{\mathcal{Q}}$  with some integration cycle  $\mathcal{Q}$  for  $\mathcal{A}$ . Since the integrand of the path integral is also a product, it follows that the path integral is the product of an integral over  $\mathcal{A}$  and an integral over  $\tilde{\mathcal{A}}$  (rather than a sum of such products, as in the more typical case (2.27)). Moreover, the integral over  $\tilde{\mathcal{A}}$  does not depend on the knot  $K$ . So the general form of  $Z_G(K; t, \tilde{t})$  is

$$Z_G(K; t, \tilde{t}) = \tilde{B}(\tilde{t}) B(K; t), \quad (2.30)$$

where only the second factor depends on the knot.

What will be  $B(K, t)$ ? Rather like the knot invariant  $Z_H(K; k)$  for the compact group  $H$ ,  $B(K, t)$  is defined by an integral of the form

$$\int_{\mathcal{Q}} D\mathcal{A} \exp(i(t/2)W(\mathcal{A})) \text{Tr}_R P \exp\left(-\oint_K \mathcal{A}\right) \quad (2.31)$$

for some integration cycle  $\mathcal{Q} \subset \hat{\mathcal{U}}_{\mathbb{C}}$ . What integration cycle shall we use?

In perturbation theory,  $Z_G(K; t, \tilde{t})$  is equal to  $Z_H(K; k)$  (with  $k = t/2 - h$ , and apart from a factor  $\tilde{B}$  that depends only on  $\tilde{t}$  and not on  $K$ ). This relation was presented as exercise 6.32 in [49]. As a perturbative statement, it simply reflects the fact that we started with a holomorphic representation whose holonomy depends on  $\mathcal{A}$  and not  $\tilde{\mathcal{A}}$ . When we compute the expectation value of  $\text{Tr}_R P \exp(-\oint_K \mathcal{A})$  in perturbation theory, it does not matter if  $\tilde{\mathcal{A}}$  is absent (the case of a compact gauge group  $H$ ) or present but decoupled from  $\mathcal{A}$  (the case of the complex gauge group  $G$ , after analytic continuation to decouple  $\tilde{\mathcal{A}}$  from  $\mathcal{A}$ ).

This relation between the  $G$  and  $H$  theories will hold as an exact statement, not just an asymptotic one, if we simply take  $\mathcal{Q}$ , the integration cycle used in constructing the

theory with complex gauge group  $G$ , to be the same integration cycle that was used in analytically continuing the theory with compact gauge group  $H$ . This is a satisfying result, but it is perhaps best understood as a definition. The original path integral (2.29) is not convergent, even for the usual real values of the coupling parameters  $s$  and  $\ell$ , if regarded as an integral over the usual real cycle  $\tilde{\mathcal{A}} = \overline{\mathcal{A}}$ . The reason for this is that the holonomy factor  $\text{Tr}_R P \exp(-\oint_K \mathcal{A})$  is exponentially growing in some directions. So the choice of  $\mathcal{Q}$  that reduces the knot invariants for  $G$  to those of  $H$  can be regarded as a definition of the knot invariants for a holomorphic representation of  $G$ , rather than something we deduce from a more primitive framework. By contrast, in the case of a unitary representation, the holonomy can be expressed as an oscillatory integral by adding more integration variables; this is explained in section 4.2. After doing so, the integral over the usual real cycle makes sense.

## 2.7 Other Real Forms

Let  $H'$  be a noncompact real form of the complex Lie group  $G = H_{\mathbb{C}}$ . (For example, if  $H = SU(2)$ ,  $H'$  might be  $SL(2, \mathbb{R})$ .) We sometimes describe this more briefly by calling  $H'$  a noncompact real form of  $H$ . Chern-Simons theory with gauge group  $H'$  is formally defined by the familiar sort of integral

$$Z_{H'}(k') = \int_{\mathcal{U}'} DA' \exp(ik'W(A')), \quad (2.32)$$

where now  $A'$  is a connection on an  $H'$  bundle  $E' \rightarrow M$  and  $\mathcal{U}'$  is the space of such connections, up to gauge transformation.

Since the complexification of  $H'$  is simply  $G = H_{\mathbb{C}}$ , the complexification of  $\mathcal{U}'$  is the familiar space  $\mathcal{U}_{\mathbb{C}}$  of complex-valued connections. Consequently,  $\mathcal{U}'$  can be viewed as an integration cycle in  $\mathcal{U}_{\mathbb{C}}$  on which the usual Chern-Simons integral is an (oscillatory) convergent integral. In this respect, it is no different from  $\mathcal{U}$ , the space of connections modulo gauge transformations for the compact Lie group  $H$ . Either  $\mathcal{U}$  or  $\mathcal{U}'$  can be regarded as a possible integration cycle in  $\mathcal{U}_{\mathbb{C}}$  for the Chern-Simons path integral. ( $\mathcal{U}$  and  $\mathcal{U}'$  are both fixed points of real involutions acting on  $\mathcal{U}_{\mathbb{C}}$ , as explained more fully in section 2.7.1.) As such, each can be expressed as a linear combination of the usual Lefschetz thimbles  $\mathcal{J}_{\sigma}$ .

Hence, the framework for analytic continuation of  $Z_{H'}(k')$  is the same as the framework for analytic continuation of  $Z_H(k)$ . We seek to expand it as a linear combination with integer coefficients of the functions  $Z_{H,\sigma}(k)$  of eqn. (2.23):

$$Z_{H'}(k') = \sum_{\sigma} n'_{\sigma} Z_{H,\sigma}(k). \quad (2.33)$$

(The coefficients  $\mathfrak{n}'_\sigma$  will have to jump in the usual way in crossing Stokes lines.) However, there is a subtlety in the relation between the parameters in  $Z_{H'}(k')$  and  $Z_H(k)$ . This relation is not  $k' = k$  but

$$k' + h' = k + h, \quad (2.34)$$

where  $h'$  is a constant<sup>6</sup> that has been computed in different ways [48, 50]. Like (2.28), which can actually be viewed as a special case, (2.34) arises from a comparison between the complexification of the integration measures used in defining  $Z_H$  and  $Z_{H'}$ . It is further discussed in section 2.7.1.

Just as in the discussion of analytic continuation of  $Z_H(k)$ , we wish to pick a linear combination of critical point cycles so that the formula (2.33) will agree with  $Z_{H'}(k')$  for  $k' \in \mathbb{Z}$ . The general analysis of this is similar to what it is for the case of compact gauge group, but again there is some simplification for the case of a knot in  $S^3$ . If we label our knot by a finite-dimensional irreducible representation of  $H'$  (as opposed to a unitary representation of  $H'$ , which would be infinite-dimensional), then it is natural to pick the same integration cycle that we used in analytically continuing  $Z_H(K, k)$ . With the identification of parameters  $k' + h' = k + h$ , this leads to  $Z_{H'}(K, k') = Z_H(K, k)$ . That relationship is already known in perturbation theory, which is insensitive to the difference between  $H$  and  $H'$  (except for a one-loop shift in the parameters [48]). That fact motivates our choice of cycle.

Our discussion in section 2.6.1 can be viewed as a special case of this, since  $G = H_{\mathbb{C}}$  is a noncompact real form of the compact Lie group  $H \times H$ . The properties of Chern-Simons theory with compact gauge group that we have used do not depend on whether the gauge group is simple, so they hold for  $H \times H$ .

### 2.7.1 Comparison Of Holomorphic Volume Forms

We will now briefly provide a context for eqn. (2.34).

The complex Lie group  $G = H_{\mathbb{C}}$  admits two real involutions that are important here: one leaves fixed  $H$  and one leaves fixed  $H'$ . Similarly, writing  $\mathfrak{h}_{\mathbb{C}}$ ,  $\mathfrak{h}$ , and  $\mathfrak{h}'$  for the Lie algebras of  $H_{\mathbb{C}}$ ,  $H$ , and  $H'$ , the space  $\mathcal{U}_{\mathbb{C}}$  of  $\mathfrak{h}_{\mathbb{C}}$ -valued connections modulo gauge transformation admits two relevant real involutions: one leaves fixed the space  $\mathcal{U}$  of  $\mathfrak{h}$ -valued connections modulo gauge transformations, and one leaves fixed the space  $\mathcal{U}'$  of  $\mathfrak{h}'$ -valued connections modulo gauge transformations.

---

<sup>6</sup>Decompose the Lie algebra of  $H'$  as  $\mathfrak{h}' = \mathfrak{k} \oplus \mathfrak{p}$ , where  $\mathfrak{k}$  is the Lie algebra of a maximal compact subgroup of  $H'$ , and  $\mathfrak{p}$  is its orthocomplement. After defining the dual Coxeter number  $h$  as the trace in  $\mathfrak{h}'$  of the square of a suitable element of  $\mathfrak{h}'$ , write  $h = h_+ + h_-$ , where  $h_+$  and  $h_-$  come from traces in  $\mathfrak{k}$  and  $\mathfrak{p}$ , respectively. Then  $h' = h_+ - h_-$ .

Accordingly, the complexification of either  $\mathcal{U}$  or  $\mathcal{U}'$  gives the same space  $\mathcal{U}_{\mathbb{C}}$ . Now let  $DA$  and  $DA'$  be the path integral measures for the theories with gauge groups  $H$  and  $H'$ , respectively. Either one of these can be analytically continued to a holomorphic volume form on  $\mathcal{U}_{\mathbb{C}}$ ; we denote these holomorphic forms simply as  $DA$  and  $DA'$ . Our discussion of analytic continuation has concerned the integration of these volume forms over various real cycles in  $\mathcal{U}_{\mathbb{C}}$  (or more precisely the integration of their pullbacks to the universal cover  $\widehat{\mathcal{U}}_{\mathbb{C}}$  over cycles in  $\widehat{\mathcal{U}}_{\mathbb{C}}$ ).

$DA$  and  $DA'$  are holomorphic volume forms on the same space, so their ratio is an invertible holomorphic function of the complexified connection  $\mathcal{A}$ . The relationship is

$$DA = DA' \exp(i(h - h')W(\mathcal{A})). \quad (2.35)$$

This relationship is computed in [48] by comparing  $\eta$ -invariants for the theories with gauge group  $H$  and  $H'$ . A different language is used there and in most of the literature. In this language, a one-loop correction (coming from the  $\eta$ -invariant) shifts the effective Chern-Simons coupling of the  $H$  theory by  $h$  and that of the  $H'$  theory by  $h'$ . The ratio of the two statements is (2.35).

The relation (2.34) is an immediate consequence of (2.35). Similarly (2.28) reflects the fact that the ratio between the holomorphic volume forms on  $\mathcal{U}_{\mathbb{C}}$  that we have called  $DA$  and  $D\mathcal{A}$  is

$$DA = D\mathcal{A} \exp(ihW(\mathcal{A})). \quad (2.36)$$

Alternatively, (2.28) is the special case of (2.34) for  $G = H_{\mathbb{C}}$  regarded as a noncompact real form of the compact Lie group  $H \times H$ .

A finite-dimensional analog of the problem of comparing such holomorphic volume forms would involve a complex manifold  $X$  with two different real structures. For example,  $X$  could be the complex  $x$ -plane, with two real structures determined by the involutions  $x \rightarrow \bar{x}$  and  $x \rightarrow -\bar{x}$ . If we set  $x = u + iv$ , the fixed set of the first involution is the  $u$  axis  $\mathcal{U}$ , with real volume form  $du$ , and the fixed point of the second is the  $v$  axis  $\mathcal{U}'$ , with real volume form  $dv$ . On the other hand, on  $X$  we have the natural holomorphic one-form  $dx = du + idv$ . The analytic continuation of  $du$  from  $\mathcal{U}$  to  $X$  is  $dx$ , and the analytic continuation of  $dv$  is  $-idx$ . So the relation between the two analytically continued volume forms is  $dv = -idx$ . In the context of the infinite-dimensional Chern-Simons path integral, such a comparison between volume forms leads to an  $\eta$  invariant, as explained in [48, 51], and this  $\eta$  invariant can be expressed in terms of the Chern-Simons function.

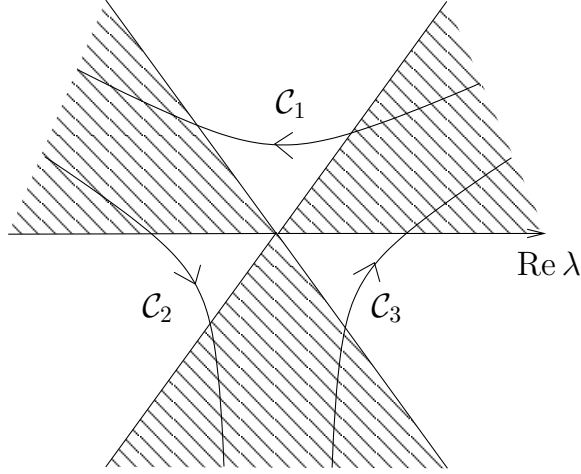


Figure 1: The complex  $\lambda$  plane, with the real axis from left to right. The hatched regions are the three “good” regions in which the Airy integrand vanishes at infinity. The contour  $\mathcal{C}_i$ ,  $i = 1, 2, 3$ , connects the  $i^{th}$  good region to the  $i + 1^{th}$ .

### 3 Finite-Dimensional Oscillatory Integrals, Morse Theory, And Steepest Descent

The goal of the present section is to describe in more detail the finite-dimensional case of the theory that was sketched in section 2.

#### 3.1 The Airy Function

We begin by examining the Airy integral, initially defined for real  $\lambda$ :

$$Z_\lambda = \int_{-\infty}^{\infty} dx \exp(i\lambda(x^3/3 - x)). \quad (3.1)$$

We want to analytically continue the integral to complex  $\lambda$ . This will require deforming the integration contour away from the real  $x$  axis, so we will regard the Airy exponent

$$\mathcal{I} = i\lambda(x^3/3 - x) \quad (3.2)$$

as a function of complex variables  $x$  and  $\lambda$ . For real  $\lambda$ , the integration is over the real axis. As  $\lambda$  varies, we will deform the integration cycle  $\mathcal{C}$  so as to vary smoothly with  $\lambda$  and in such a way that the integral always converges.



In the case of the Airy function, it is not hard to see by hand how to do this. We divide the  $x$ -plane into “good” regions in which the imaginary part of  $\lambda x^3$  is positive and “bad” regions in which it is negative. For positive  $\lambda$ , this is shown in fig. 1. The Airy integrand is exponentially small in the good regions. The original oscillatory integral over the real  $x$  axis is an integral over the boundary of two good regions. We can deform the contour slightly by giving  $x$  a positive imaginary part so that it will connect two good regions. Now as we vary  $\lambda$  (keeping away from  $\lambda = 0$ ), the good regions rotate continuously to the left or the right, and we simply rotate the integration contour  $\mathcal{C}$  as  $\lambda$  varies so that  $\mathcal{C}$  always connects two good regions. This gives an analytic continuation of  $Z_\lambda$  as a function of  $\lambda$ , but it also makes it clear that there is a monodromy around  $\lambda = \infty$  (or 0). In fact, as  $\lambda$  changes in phase by  $2\pi$ , the three good regions are cyclically permuted. So  $Z_\lambda$  has a monodromy of order 3.

Because the contour  $\mathcal{C}$  does undergo a non-trivial monodromy as  $\lambda$  varies, it is useful to consider all possible contours in this problem. Denote the three good regions as  $\mathcal{V}_i$ ,  $i = 1, 2, 3$ , as in the figure. For each  $i$ , let  $\mathcal{C}_i$  be a contour connecting region  $\mathcal{V}_i$  to region  $\mathcal{V}_{i+1}$ . The integrals  $Z_{i,\lambda} = \int_{\mathcal{C}_i} dx \exp(i\lambda(x^3/3 - x))$  converge for all  $i$  and  $\lambda$ , and any reasonable contour along which one might integrate the Airy function  $\exp(i\lambda(x^3/3 - x))$  (or the product of this function and one with slower growth at infinity) is equivalent to a linear combination of the  $\mathcal{C}_i$ . However, the integrals  $Z_{i,\lambda}$  are not independent. They obey

$$Z_{1,\lambda} + Z_{2,\lambda} + Z_{3,\lambda} = 0. \quad (3.3)$$

This reflects the fact that the sum of contours  $\mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3$  can be deformed to zero.

### 3.1.1 Morse Theory

The Airy function is simple enough that we do not really need a more sophisticated theory. Nevertheless, we will use this simple example as a practice case for describing a more sophisticated approach that is useful in generalizations. To formalize the sense in which  $\mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3 = 0$ , it is useful to regard the  $\mathcal{C}_i$  as relative homology cycles. Let  $X$  denote the complex  $x$ -plane, and for real  $U$ , let  $X_U$  be the part of  $X$  with  $\text{Re}(i\lambda(x^3/3 - x)) \leq U$ . Any reasonable integration cycle  $\mathcal{C}$  for the Airy integral should represent an element of the relative homology  $H_1(X, X_{-T}; \mathbb{Z})$  for very large  $T$ . In other words,  $\mathcal{C}$  is a closed cycle, and though it may not be compact, its infinite ends should lie in  $X_{-T}$ . Differently put,  $\text{Re}(i\lambda(x^3/3 - x))$  must tend to  $-\infty$  at infinity in  $\mathcal{C}$ .

The  $\mathcal{C}_i$  have this property, and represent elements of the relative homology. Indeed, they generate this relative homology, subject to the relation

$$\mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3 = 0. \quad (3.4)$$

The fact that the  $\mathcal{C}_i$  generates the relative homology is fairly obvious geometrically in this simple example.  $X$  is contractible, but  $X_{-T}$  is the union of three disjoint pieces. A one-cycle that is non-trivial in  $H_1(X, X_{-T}; \mathbb{Z})$  must connect distinct pieces of  $X_{-T}$ . The relation (3.4) is also clear geometrically.

The integration cycle that we started with for real  $\lambda$  – the real  $x$ -axis, which we will call  $\mathcal{C}_{\mathbb{R}}$  – is not quite an element of the relative homology, because at large  $|x|$  it lies on the boundary of, rather than in the interior of, the good regions of fig. (1). However,  $\mathcal{C}_{\mathbb{R}}$  becomes an element of the relative homology once we displace it slightly by giving  $x$  a small imaginary part  $\epsilon$  with the same sign as  $\lambda$ . This multiplies the Airy integral by a convergence factor  $\exp(-\lambda\epsilon x^2)$ . The displaced cycle therefore represents an element of the relative homology. And indeed since the displaced cycle connects two of the good regions at infinity, it is equivalent to one of the  $\mathcal{C}_i$ .

For generalizations, we need a more powerful way to analyze the relative homology. We consider the real part of the Airy exponent

$$h = \text{Re}(\mathcal{I}) \quad (3.5)$$

as a Morse function. A Morse function is simply a real-valued function whose critical points are nondegenerate. The critical points are the points at which all first derivatives of  $h$  vanish, and nondegeneracy of a critical point means that the matrix of second derivatives is invertible at that point. The number of negative eigenvalues of this matrix is called the Morse index of the critical point.

In general, because of the Cauchy-Riemann equations, the critical points of the real part of a holomorphic function  $\mathcal{I}$  are the same as the critical points of  $\mathcal{I}$ . So in this case, there are two critical points which we call  $p_{\pm}$ , located at  $x = \pm 1$ . The values of  $\mathcal{I}$  and  $h$  at  $p_{\pm}$  are

$$\mathcal{I}_{\pm} = \mp \frac{2i\lambda}{3} \quad (3.6)$$

and

$$h_{\pm} = \pm \frac{2\text{Im}\lambda}{3}. \quad (3.7)$$

The critical points are nondegenerate, since  $\partial^2 \mathcal{I} / \partial x^2 \neq 0$  at  $x = \pm 1$ . In general, in a complex manifold of complex dimension  $n$ , if  $h$  is a Morse function that is the real part of a holomorphic function  $\mathcal{I}$ , its isolated and nondegenerate critical points are all of index  $n$ . (The general form of  $\mathcal{I}$  near an isolated nondegenerate critical point is  $\mathcal{I} = c + \sum_{i=1}^n z_i^2$ , with a constant  $c$  and local complex coordinates  $z_i$ ; if  $z_i = u_i + iv_i$  then  $h = \text{Re } c + \sum_{i=1}^n (u_i^2 - v_i^2)$  and the matrix of its second derivatives is diagonal with  $n$  negative eigenvalues.) So in our case, the critical points have index 1.

In general, any Morse function  $h$  on a manifold  $X$  can be used to compute upper bounds on the ranks of the real homology (or cohomology) groups of  $X$ : the rank of the  $q$ -dimensional

homology of  $X$  is at most the number of critical points of  $h$  of Morse index  $q$ . If the differences between the indices of distinct critical points of  $h$  are never equal<sup>7</sup> to  $\pm 1$ , then  $h$  is a perfect Morse function and the ranks of the homology groups are equal to the upper bounds just stated. In fact, in this case even the integral homology can be described via the critical points. Our Morse function is perfect, since both critical points have index 1.

If  $X$  is a compact manifold,  $h$  is automatically bounded above and below, and the critical points of a perfect Morse function determine the ordinary homology of  $X$ . In our case,  $X$  is not compact, and we are dealing with a Morse function  $h$  that is unbounded above and below. In such a case, the critical points of  $h$  determine relative homology groups, namely the homology groups  $H_k(X, X_{-T})$ , where  $T$  is a large constant. In our example,  $h$  is a Morse function with two critical points both of index 1, so  $H_1(X, X_{-T})$  is of rank 2, and the other relative homology groups vanish. (Without using Morse theory, we would observe that  $H_0(X, X_{-T}) = 0$  since any zero-cycle is a sum of points, which can be deformed into  $X_{-T}$ . And  $H_2(X, X_{-T}) = 0$  since  $X$  has non-compact ends even after collapsing  $X_{-T}$  to a point.)

For a generic perfect Morse function  $h$ , Morse theory also gives a recipe for constructing a relative one-cycle  $\mathcal{J}_\sigma$  attached to each critical point, and representing an element of the appropriate relative homology group. Moreover, these cycles generate the relative homology with integer coefficients.

In general, on any manifold  $X$  with real coordinates  $u^i$ , pick a Riemannian metric  $g_{ij}$  and consider the downward flow or gradient flow equations

$$\frac{du^i}{dt} = -g^{ij} \frac{\partial h}{\partial u^j}. \quad (3.8)$$

These are called downward flow equations, because except for a trivial solution that sits at a critical point for all  $t$ ,  $h$  is always strictly decreasing along a flow:

$$\frac{dh}{dt} = \sum_i \frac{\partial h}{\partial u^i} \frac{du^i}{dt} = - \sum_i \left( \frac{\partial h}{\partial u^i} \right)^2. \quad (3.9)$$

An important property of the flow equation is that if  $u^i(t)$  equals a critical point at some  $t$ , then the flow equation implies that  $u^i(t)$  is constant for all  $t$ . So a nonconstant flow can only reach a critical point at  $t = \pm\infty$ .

Let  $p$  be a nondegenerate critical point of  $h$ , and consider the downward flow equations on the half-line  $(-\infty, 0]$  with the boundary condition that  $u^i(t)$  approaches  $p$  for  $t \rightarrow -\infty$ . If  $p$  has index  $k$ , the moduli space  $\mathcal{J}_p$  of such solutions is a  $k$ -dimensional manifold, since there are  $k$  independent directions of downward flow from the critical point  $p$ . We think

---

<sup>7</sup>This condition ensures that in computing the homology of  $X$ , there are no cancellations between critical points of adjacent index.

of  $\mathcal{J}_p$  as a submanifold of  $X$ , by mapping a downward flow line  $u^i(t)$  to the corresponding point  $u^i(0) \in X$ . (This gives an embedding of  $\mathcal{J}_p$  in  $X$  since, as the downward flow equation is first order in time, a flow is uniquely determined by its value at  $t = 0$ .) An equivalent definition of  $\mathcal{J}_p$  is that it is the submanifold of  $X$  consisting of points that can be reached at any  $t$  by a flow that starts at  $p$  at  $t = -\infty$ . Given a flow line  $u^i(t)$  that reaches a point  $x \in X$  at  $t = t^*$ , the flow line  $u^i(t - t^*)$  arrives at  $x$  at  $t = 0$ . A flow line defined on the full line  $(-\infty, \infty)$  will be called a complete flow line.

Once we pick an orientation of  $\mathcal{J}_p$ ,  $\mathcal{J}_p$  will define a cycle<sup>8</sup> in the relative homology  $H_1(X, X_{-T})$  if it is closed, meaning that any sequence of points in  $\mathcal{J}_p$  has a subsequence that either converges or tends to  $h = -\infty$ . This fails precisely if there is a complete flow line  $l$  that starts at  $p$  at  $t = -\infty$  and ends at another critical point  $q$  at  $t = +\infty$ . In that case,  $\mathcal{J}_p$  is not closed, since  $l$  is contained in  $\mathcal{J}_p$ , but a sequence of points in  $l$  can converge to  $q$ , which is not contained in  $\mathcal{J}_p$ .

We will determine shortly a necessary condition for a flow line to connect two distinct critical points  $p$  and  $q$ . When this condition is not satisfied,  $\mathcal{J}_p$  defines an element of the relative homology  $H_k(X, X_{-T})$ .  $\mathcal{J}_p$  is known as a Lefschetz thimble.

In our problem, the Morse function is the real part of a holomorphic function  $\mathcal{I}$ . In such a case, if we pick a Kahler metric  $ds^2 = |dx|^2$ , the flow equation can be written

$$\frac{dx}{dt} = -\frac{\partial \bar{\mathcal{I}}}{\partial \bar{x}}, \quad \frac{d\bar{x}}{dt} = -\frac{\partial \mathcal{I}}{\partial x}. \quad (3.10)$$

Using the chain rule and (3.10), we find

$$\frac{d \operatorname{Im} \mathcal{I}}{dt} = \frac{1}{2i} \frac{d(\mathcal{I} - \bar{\mathcal{I}})}{dt} = \frac{1}{2i} \left( \frac{\partial \mathcal{I}}{\partial x} \frac{dx}{dt} - \frac{\partial \bar{\mathcal{I}}}{\partial \bar{x}} \frac{d\bar{x}}{dt} \right) = 0. \quad (3.11)$$

So  $\operatorname{Im} \mathcal{I}$  is conserved along a flow.

A more conceptual explanation of the existence of this conserved quantity is as follows. Associated to the Kahler metric  $|dx|^2$  is the Kahler form  $\omega = -i dx \wedge d\bar{x}$ . Viewing  $\omega$  as a symplectic form on the  $x$ -plane, the gradient flow with respect to the Morse function  $\operatorname{Re} \mathcal{I}$  is the same as the Hamiltonian flow with Hamiltonian  $\operatorname{Im} \mathcal{I}$ . Naturally,  $\operatorname{Im} \mathcal{I}$  is conserved in this flow.

From (3.6), we see that in the Airy case, the values of  $\operatorname{Im} \mathcal{I}$  at the critical points are  $\operatorname{Im} \mathcal{I}_{\pm} = \mp(2/3)\operatorname{Re} \lambda$ . These are equal if and only if  $\lambda$  is purely imaginary. Hence a flow

---

<sup>8</sup>Our definition of  $\mathcal{J}_p$  is slightly different from the definition usually used in Morse theory on a compact manifold  $M$ . In such a case, unless  $p$  is a local minimum of  $h$ , there are always complete flow lines starting at  $p$  and ending at some “lower” critical point, and one takes the closure of  $\mathcal{J}_p$  to define a Morse theory cycle. In the situation we consider, as is explained shortly, there generically are no complete flow lines between critical points, and it is most convenient to define  $\mathcal{J}_p$  as we have done.

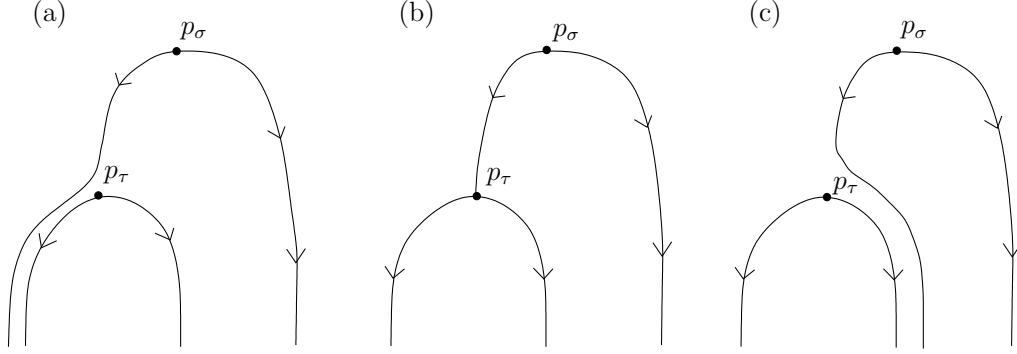


Figure 2: This figure illustrates the behavior of flow lines when one crosses a Stokes ray. Sketched are the downward flows from two critical points  $p_\sigma$  and  $p_\tau$ . The lines are sketched as flowing downward, in the direction of smaller  $h$  (though it is not literally true that  $h$  corresponds to the height in the figure, as this function has no critical points). The behavior at a Stokes ray is depicted in (b); there is a downward flow from  $p_\sigma$  to  $p_\tau$ . The flows “before” and “after” crossing the Stokes ray are depicted in (a) and (c). The downward flowing cycle  $\mathcal{J}_\tau$  from the lower critical point is unaffected by the Stokes ray. As for the cycle  $\mathcal{J}_\sigma$  defined by flow from the upper critical point, it is ill-defined in (b) and jumps by  $\mathcal{J}_\sigma \rightarrow \mathcal{J}_\sigma + \mathcal{J}_\tau$  between (a) and (c).

can connect the two critical points at  $x = \pm 1$  only if  $\lambda$  is imaginary. Conversely, if  $\lambda$  is imaginary, there is indeed a flow that connects the two critical points (the flow line is simply the part of the real axis with  $-1 < x < 1$ ). Since we always assume that  $\lambda \neq 0$ , the locus of imaginary  $\lambda$  is not connected but consists of two open rays, differing by the sign of  $\text{Im } \lambda$ . We refer to these as Stokes rays.

### 3.1.2 Crossing A Stokes Ray

Away from the Stokes rays, a downward flow that starts at one critical point cannot end at the other; instead it always flows to  $h = -\infty$ . So away from the Stokes rays, the downward flow procedure attaches to each critical point  $p_+$  or  $p_-$  a relative homology one-cycle  $\mathcal{J}_\pm$ . Let us see what happens to the Lefschetz thimbles  $\mathcal{J}_\pm$  in crossing a Stokes ray.

From (3.6),  $\mathcal{I}_+$  and  $\mathcal{I}_-$  are always unequal (since we always assume that  $\lambda \neq 0$ ). On a Stokes ray  $\text{Im } \mathcal{I}_+ = \text{Im } \mathcal{I}_-$ , so necessarily  $\text{Re } \mathcal{I}_+ \neq \text{Re } \mathcal{I}_-$ , that is  $h_+ \neq h_-$ . In fact,  $h_+ > h_-$  if  $\text{Im } \lambda > 0$ , and otherwise  $h_- > h_+$ . As we approach a Stokes ray, let us write  $q$  and  $q'$  for the critical points with the larger and smaller values of  $h$ , and  $\mathcal{J}$  and  $\mathcal{J}'$  for the corresponding one-cycles. Even when we reach and cross the Stokes ray, nothing happens to  $\mathcal{J}'$ , because downward trajectories that start at  $q'$  can only flow to  $h = -\infty$ . But a trajectory that starts at  $q$  may flow down to  $q'$ . (As we noted above, in the case of the Airy function, there is such

a trajectory, namely the segment of the real axis connecting  $q$  and  $q'$ .) When there is such a flow line, the definition of the cycle  $\mathcal{J}$  attached to  $q$  fails. Since  $\mathcal{J}$  is potentially ill-defined when  $\lambda$  is on a Stokes ray, it might jump in crossing a Stokes ray, and that is actually what happens. How this happens is sketched in Fig. 2. The jump is  $\mathcal{J} \rightarrow \mathcal{J} \pm \mathcal{J}'$ , where the sign depends on the orientations of the two cycles and the direction in which  $\lambda$  crosses the Stokes ray. The change in the pair  $\mathcal{J}, \mathcal{J}'$  can thus be written

$$\begin{pmatrix} \mathcal{J} \\ \mathcal{J}' \end{pmatrix} \rightarrow \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathcal{J} \\ \mathcal{J}' \end{pmatrix}. \quad (3.12)$$

In section 3.1.4, we explain in another way why only  $\mathcal{J}$  and not  $\mathcal{J}'$  changes in crossing the Stokes ray.

Let us compute the monodromy of the cycles  $\mathcal{J}_+$  and  $\mathcal{J}_-$  around  $\lambda = \infty$  (or 0). In crossing the Stokes ray with  $\text{Im } \lambda > 0$ , the monodromy is strictly upper triangular:

$$U_+ = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (3.13)$$

On the Stokes ray with  $\text{Im } \lambda < 0$ , the monodromy is strictly lower triangular:

$$U_- = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}. \quad (3.14)$$

(If the cycles  $\mathcal{J}_\pm$  are oriented so that the upper right matrix element in (3.13) is +1, then the lower left matrix element in (3.14) is -1.) The monodromy around  $\lambda = \infty$  is not simply the product  $U_- U_+$ , as there is also a minus sign involving the orientation of the cycles  $\mathcal{J}_\pm$ . Indeed, the function  $\mathcal{I}$  can be approximated near  $x = \pm 1$  by  $\mp 2i\lambda/3 + i\lambda(x \mp 1)^2$ , and an approximation to the downward flowing cycles  $\mathcal{J}_\pm$  is given by  $x = \pm 1 + t(-i\lambda)^{-1/2}$ , with real  $t$ . To orient the cycles, we need to pick a sign of  $(-i\lambda)^{-1/2}$ , and under monodromy around  $\lambda = \infty$ , this sign is reversed. So the monodromy of the pair  $\begin{pmatrix} \mathcal{J}_+ \\ \mathcal{J}_- \end{pmatrix}$  around  $\lambda = \infty$  is

$$M = -U_- U_+ = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}. \quad (3.15)$$

In particular, the monodromy obeys  $M^3 = 1$ . The fact that the monodromy is of order 3 is clear without Morse theory: under monodromy around  $\lambda = \infty$ , the three good regions in Fig. 1 undergo a cyclic permutation of order 3. What we have gained from Morse theory is a powerful framework for generalizations.

In this analysis, the dependence of the orientation of the  $\mathcal{J}_\pm$  on a sign of  $\sqrt{-i\lambda}$  is inelegant, and the resulting jumping in the sign of the  $\mathbf{n}_\sigma$  is unthematic. This can be avoided by multiplying the original definition of the Airy integral by  $\sqrt{-i\lambda}$ . The product of this function times the orientation of  $\mathcal{J}_\pm$  has no jumping in sign. With this factor included, the jumping of  $\mathbf{n}_\sigma$  occurs only in crossing Stokes lines, and  $M$  has the opposite sign to what is given in (3.15) and obeys  $M^3 = -1$ .

### 3.1.3 Asymptotic Behavior Of Integrals Over Lefschetz Thimbles

The cycles  $\mathcal{J}_\pm$  are defined in such a way that, for any  $\lambda$ , the integrals

$$Z_{\pm,\lambda} = \int_{\mathcal{J}_\pm} dx \exp(\mathcal{I}) \quad (3.16)$$

always converge. Indeed, we have arranged so that  $\text{Re } \mathcal{I} \rightarrow -\infty$  at infinity along  $\mathcal{J}_\pm$ .

Not only are the integrals  $Z_{\pm,\lambda}$  always convergent, but it is straightforward to determine their asymptotic behavior as  $\lambda \rightarrow \infty$  in a fixed direction in the complex plane. The maximum of  $h = \text{Re } \mathcal{I}$  on the cycle  $\mathcal{J}_\pm$  is precisely at the critical point  $x = \pm 1$ . Moreover, the phase  $\text{Im } \mathcal{I}$  is stationary at  $x = \pm 1$ , since those are critical points of  $\mathcal{I}$  (and in fact the phase  $\text{Im } \mathcal{I}$  is entirely constant along the cycles  $\mathcal{J}_\pm$ ). So the asymptotic behavior of  $Z_{\pm,\lambda}$  is dominated by a contribution from the critical point, giving

$$Z_{\pm,\lambda} \sim \exp(\mathcal{I}_\pm) (-i\lambda)^{-1/2} \sum_{t=0}^{\infty} b_t \lambda^{-t}, \quad (3.17)$$

where  $\mathcal{I}_\pm = \mp 2i\lambda/3$  and the coefficients  $b_t$  in the asymptotic series can be computed by expanding around the critical point. If we include a factor of  $(-i\lambda)^{1/2}$  in the definition of the Airy integral, as suggested at the end of section 3.1.1, the factor of  $(-i\lambda)^{-1/2}$ , which comes from a Gaussian integral, would be canceled.

### 3.1.4 Sum Of Asymptotic Series

An integral  $Z_{\mathcal{C},\lambda}$  over a general cycle  $\mathcal{C} = \mathbf{n}_+ \mathcal{J}_+ + \mathbf{n}_- \mathcal{J}_-$  can be expressed asymptotically as a sum of two asymptotic expansions:

$$Z_{\mathcal{C},\lambda} \sim \mathbf{n}_+ \exp(\mathcal{I}_+) (-i\lambda)^{-1/2} \sum_{t=0}^{\infty} b_t \lambda^{-t} + \mathbf{n}_- \exp(\mathcal{I}_-) (-i\lambda)^{-1/2} \sum_{t=0}^{\infty} c_t \lambda^{-t}. \quad (3.18)$$

However, generically one of these series suffices, since in a generic direction in the complex  $\lambda$  plane,  $h_+$  and  $h_-$  are unequal and one of the two series is exponentially larger than the other. If  $h_+ = h_-$  (which happens for real  $\lambda$ ), the two series compete and must both be included, and if  $\lambda$  varies so that  $h_+ - h_-$  changes sign, the two series exchange dominance.

On a Stokes ray,  $h_+$  and  $h_-$  are definitely unequal, so one series in (3.18) is exponentially larger than the other. Consider for definiteness a Stokes ray with  $h_+ > h_-$ . The integral in (3.18) must vary holomorphically with  $\lambda$ , so the coefficient  $\mathbf{n}_+$  of the dominant exponential series must be continuous. But, since the two series are only asymptotic series that leave

exponentially small corrections undetermined, the expansion (3.18) in a region with  $h_+ > h_-$  is not powerful enough to ensure that  $\mathbf{n}_-$  is continuous.

Let us compare this expectation to (3.12), which in the present notation says that the jumping of  $\mathcal{J}_+$  and  $\mathcal{J}_-$  is

$$\begin{pmatrix} \mathcal{J}_+ \\ \mathcal{J}_- \end{pmatrix} \rightarrow \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathcal{J}_+ \\ \mathcal{J}_- \end{pmatrix}. \quad (3.19)$$

In order for  $\mathbf{n}_+\mathcal{J}_+ + \mathbf{n}_-\mathcal{J}_-$  to be continuous, the  $\mathbf{n}_\pm$  transform by

$$\begin{pmatrix} \mathbf{n}_+ & \mathbf{n}_- \end{pmatrix} \rightarrow \begin{pmatrix} \mathbf{n}_+ & \mathbf{n}_- \end{pmatrix} \begin{pmatrix} 1 & \mp 1 \\ 0 & 1 \end{pmatrix}. \quad (3.20)$$

In particular, as expected,  $\mathbf{n}_+$  is constant and only  $\mathbf{n}_-$  jumps.

An equivalent way to say some of the same things is to observe that since the integral over  $\mathcal{J}_-$  is exponentially smaller than that over  $\mathcal{J}_+$ ,  $\mathcal{J}_-$  cannot jump by a multiple of  $\mathcal{J}_+$  in crossing the Stokes ray, but  $\mathcal{J}_+$  can jump by a multiple of  $\mathcal{J}_-$ .

### 3.1.5 Expressing The Real Cycle In Terms Of Lefschetz Thimbles

The function  $Z_\lambda$  whose analytic continuation we actually want to understand is defined for real  $\lambda$  by integration over the real cycle  $\mathcal{C}_\mathbb{R}$ . So to use the above results, we have to express  $\mathcal{C}_\mathbb{R}$  as a linear combination of the  $\mathcal{J}_\pm$ . The coefficients will be integers, since the  $\mathcal{J}_\pm$  are a basis of the relative homology with integer coefficients. For the Airy integral, we can find the right coefficients by inspection, but we want to describe a more general method.

On a compact Riemann surface  $X$ , given a basis  $\mathcal{J}_\pm$  of  $H_1(X)$ , to express another class  $\mathcal{C} \in H_1(X)$  in the form  $\mathcal{C} = \mathbf{n}_+\mathcal{J}_+ + \mathbf{n}_-\mathcal{J}_-$ , we would find the coefficients  $\mathbf{n}_+$  and  $\mathbf{n}_-$  via intersection pairings. In the present case, we cannot do precisely this, because there is no intersection pairing on the relative homology  $H_1(X, X_{-T})$ . (The intersection of two one-cycles on the two-manifold  $X$  should be a zero-cycle or sum of points; but a point vanishes in the relative homology  $H_0(X, X_{-T})$ , since it can be deformed into  $X_{-T}$ .)

There is a good substitute, however. Indeed, Morse theory tells us what are the duals of the downward flowing cycles from the critical points: they are the upward flowing cycles. The upward flowing cycle attached to a critical point  $p$  is defined by reversing the sign in the flow equation (3.8). That is, one considers solutions of the upward flowing equation

$$\frac{du^i}{dt} = +g^{ij} \frac{\partial h}{\partial u^j} \quad (3.21)$$

on the half-line  $(-\infty, 0]$ , again requiring that  $u^i(t)$  approaches  $p$  for  $t \rightarrow -\infty$ . The space of possible values of  $u^i(t)$  at  $t = 0$  now defines the upward flowing cycle  $\mathcal{K}_p$  attached to  $p$ . It is



an  $n - k$ -dimensional cycle if  $p$  has index  $k$  (so that the Morse function  $-h$  has index  $n - k$  at  $p$ ).

The upward flowing cycles take values in a different relative homology. For a real number  $T$ , define  $X^T$  as the part of the complex  $x$ -plane with  $h \geq T$ . The upward flowing cycle from a critical point  $p$  of the type we are considering has its infinite ends contained in  $X^T$  for large positive  $T$  (as long as we keep away from Stokes rays). So it defines an element of the relative homology  $H_k(X, X^T)$ .

For  $r$  and  $s$  of complementary dimensions, there is a natural intersection pairing between  $H_r(X, X_{-T})$  and  $H_s(X, X^T)$ . This pairing is easily evaluated for the case of interest to us – a perfect Morse function with no flows between distinct critical points. The absence of such flows means that if  $p_\sigma$  and  $p_\tau$  are distinct critical points, the corresponding cycles  $\mathcal{J}_\sigma$  and  $\mathcal{K}_\tau$  do not intersect – such an intersection would be a downward flow from  $p_\sigma$  to  $p_\tau$ . On the other hand, for any one critical point  $p_\sigma$ , the downward cycle  $\mathcal{J}_\sigma$  and the upward cycle  $\mathcal{K}_\sigma$  have precisely one point of intersection – the trivial constant flow, starting and ending at  $p_\sigma$ . With suitable relative orientations of downwards and upward cycles, the trivial flow contributes  $+1$  to the diagonal intersection pairing  $\langle \mathcal{J}_\sigma, \mathcal{K}_\sigma \rangle$ . So the intersection pairings are

$$\langle \mathcal{J}_\sigma, \mathcal{K}_\tau \rangle = \delta_{\sigma\tau}. \quad (3.22)$$

Given any element  $\mathcal{C}$  of the downward flowing relative homology, we now can give a formula to determine the coefficients in the expansion  $\mathcal{C} = \sum_\sigma \mathbf{n}_\sigma \mathcal{J}_\sigma$ . By intersecting with  $\mathcal{K}_\tau$ , we find simply

$$\mathbf{n}_\tau = \langle \mathcal{C}, \mathcal{K}_\tau \rangle \quad (3.23)$$

for all  $\tau$ .

### 3.1.6 The Asymptotic Behavior Of The Airy Function

We will now give some examples of this procedure. Return to the original Airy function

$$Z_\lambda = \int_{\mathcal{C}_\mathbb{R}} dx \exp(i\lambda(x^3/3 - x)) \quad (3.24)$$

with  $\mathcal{C}_\mathbb{R}$  being the real axis. The critical points  $p_\pm$  at  $x = \pm 1$  both lie on  $\mathcal{C}_\mathbb{R}$ . A nontrivial upward flow starting at a critical point will end at a point with a strictly positive value of  $h = \text{Re}\mathcal{I}$ ; such a point will not lie on  $\mathcal{C}_\mathbb{R}$ , since  $h$  vanishes identically on  $\mathcal{C}_\mathbb{R}$ . Hence the intersection of  $\mathcal{C}_\mathbb{R}$  with either upward flowing cycle  $\mathcal{K}_\pm$  consists precisely of the critical point  $p_\pm$ . Modulo a judicious choice of orientations, this tells us the coefficients when we express the real cycle  $\mathcal{C}_\mathbb{R}$  in terms of Lefschetz thimbles associated to critical points:

$$\mathcal{C}_\mathbb{R} = \mathcal{J}_+ + \mathcal{J}_-. \quad (3.25)$$

This formula expresses the integration cycle in terms of Lefschetz thimbles when  $\lambda$  is real. But when  $\lambda$  is not real, we must integrate over a more general cycle  $\mathcal{C}$ . As  $\lambda$  varies in the complex plane, we vary the cycle  $\mathcal{C}$  smoothly so that it always connects one good region at infinity to another. On the other hand, the cycles  $\mathcal{J}_+$  and  $\mathcal{J}_-$  have jumps described in eqns. (3.13), (3.14) and a jump in sign involving the orientations. Allowing for these jumps, we get, after analytic continuation from real  $\lambda$  to a wedge-shaped region of the complex  $\lambda$  plane that does not contain Stokes rays, a formula

$$\mathcal{C} = \mathbf{n}_+ \mathcal{J}_+ + \mathbf{n}_- \mathcal{J}_-. \quad (3.26)$$

(Instead of keeping track of Stokes rays and orientations, one can in this particular problem let  $\mathcal{C}$  evolve continuously so that the Airy integral remains convergent, and then directly use (3.23) to determine the coefficients in any wedge-shaped region.) Hence in each wedge-shaped region, the Airy integral can be expressed as

$$Z_\lambda = \mathbf{n}_+ Z_{+, \lambda} + \mathbf{n}_- Z_{-, \lambda}. \quad (3.27)$$

Together with (3.17), this determines the asymptotic behavior of  $Z_\lambda$  in each wedge-shaped region. In particular,  $Z_\lambda$  grows exponentially at infinity if and only if the expansion (3.26) has a nonzero coefficient for a critical point at which  $h > 0$ .

Let us consider a special case in which there is *not* exponential growth at infinity. This is the Airy function with a reversed sign for the linear term in the exponent,

$$\tilde{Z}_\lambda = \int_{\mathcal{C}_\mathbb{R}} dx \exp(i\lambda(x^3/3 + x)), \quad (3.28)$$

with  $\lambda$  real and with  $\mathcal{C}_\mathbb{R}$  again equal to the real axis. This integral is actually a special case of what one gets by analytic continuation of the original Airy integral, since<sup>9</sup> (3.24) is mapped to (3.28) by  $\lambda \rightarrow i\lambda$ ,  $x \rightarrow ix$ . However, we will simply treat this slightly modified example directly according to the general recipe.

The critical points are now located at  $x = \pm i$ . We denote these points as  $\tilde{p}_\pm$ . The value of  $\tilde{\mathcal{I}} = i\lambda(x^3/3 + x)$  at  $\tilde{p}_\pm$  is now  $\mp 2\lambda/3$ , so for either sign of  $\lambda$ , the function  $\tilde{h} = \text{Re } \tilde{\mathcal{I}}$  is positive at one critical point and negative at the other critical point. An upward flow from a critical point with  $\tilde{h} > 0$  will never meet the real axis  $\mathcal{C}_\mathbb{R}$ , since  $\tilde{h} = 0$  on  $\mathcal{C}_\mathbb{R}$ . However, from the critical point with  $\tilde{h} < 0$ , there is an upward flow to the real axis.<sup>10</sup> Write  $q^+$  and

---

<sup>9</sup>To be more exact, to map the original Airy function to (3.28) by analytic continuation, one must increase the argument of  $\lambda$  by  $2\pi(n + 1/4)$ , where the integer  $n$  is chosen so that the real cycle of the original Airy function maps back to the real cycle in (3.28).

<sup>10</sup> For example, if  $\lambda$  is real, the part of the imaginary axis between  $x = i$  and  $x = -i$  is a flow line that intersects the real axis. In one direction or the other, this is an upward flow from the critical point with negative  $\tilde{h}$ . If we give  $\lambda$  a small imaginary part to get away from the Stokes ray, the upward flow from that same critical point still intersects the real axis (it then continues to  $\tilde{h} = +\infty$  rather than to the second critical point).

$q^-$  for the critical points at which  $\tilde{h}$  is positive or negative, and  $\tilde{\mathcal{J}}_{\pm}$  for the corresponding Lefschetz thimbles. Then the expression for  $\mathcal{C}_{\mathbb{R}}$  in terms of Lefschetz thimbles is simply

$$\mathcal{C}_{\mathbb{R}} = \tilde{\mathcal{J}}_{-}. \quad (3.29)$$

The asymptotic behavior of  $\tilde{Z}_{\lambda}$  for large real  $\lambda$  is hence

$$\tilde{Z}_{\lambda} \sim \frac{\exp\left(\tilde{\mathcal{I}}(q^-)\right)}{\lambda^{1/2}} \sum_{t=0}^{\infty} a_t \lambda^{-t}. \quad (3.30)$$

In particular,  $\tilde{Z}_{\lambda}$  vanishes exponentially for large real  $\lambda$ , reflecting the fact that there are no critical points on the real  $x$ -axis.

Actually, in this example, the Stokes rays are the positive and negative real  $\lambda$  axis. For real  $\lambda$ , there is a flow line between the two critical points, as noted in footnote 10. Hence the cycles  $\tilde{\mathcal{J}}_{\pm}$  jump in crossing the real  $\lambda$  axis. However, the jumping only affects the cycle  $\tilde{\mathcal{J}}_{+}$  attached to the critical point with  $\tilde{h} > 0$ . This cycle changes by  $\tilde{\mathcal{J}}_{+} \rightarrow \tilde{\mathcal{J}}_{+} \pm \tilde{\mathcal{J}}_{-}$ . The cycle  $\tilde{\mathcal{J}}_{-}$  is unchanged in crossing the Stokes ray, and hence there is no problem with the formula (3.29).

## 3.2 Generalization To Dimension $n$

In our analysis of the Airy function, we introduced much more machinery than was really needed. The advantage of having done so is that the generalization to an oscillatory integral of higher dimension is now straightforward.

We consider a real-valued polynomial function  $f(x^1, \dots, x^n)$  of  $n$  real variables, and we assume that  $f$  is sufficiently generic that (even after analytic continuation to complex values of the  $x^i$ ) it has only finitely many critical points, which are all nondegenerate. We want to study the analytic continuation of the oscillatory integral<sup>11</sup>

$$Q_{\lambda} = (-i\lambda)^{n/2} \int_{\mathcal{C}_{\mathbb{R}}} d^n x \exp\left(i\lambda f(x^1, \dots, x^n)\right). \quad (3.31)$$

The integration is over the cycle  $\mathcal{C}_{\mathbb{R}}$  defined by taking  $x^1, \dots, x^n$  to be real. We wish to analytically continue the integral to (nonzero) complex  $\lambda$ , and to determine the asymptotic behavior of the analytically continued integral. Depending on  $f$ , it may be difficult to see how to vary the integration cycle  $\mathcal{C}$  as  $\lambda$  is varied so that the integral remains convergent.

---

<sup>11</sup>Including the elementary prefactor  $(-i\lambda)^{n/2}$  avoids sign changes in the orientations of the Lefschetz thimbles, as explained at the end of section 3.1.1. This lets us concentrate on the more interesting Stokes phenomena.

So it is difficult in general to imitate the direct approach to the Airy function by rotating  $\mathcal{C}$  as  $\lambda$  varies. But the approach via Morse theory can be straightforwardly generalized.

As in the one-variable case, we regard the function  $\mathcal{I} = i\lambda f(x^1, \dots, x^n)$  as a function of complex variables, and for fixed complex  $\lambda$ , we let  $\Sigma$  be the set of its critical points. If  $p_\sigma$ ,  $\sigma \in \Sigma$ , is a critical point, we write  $\mathcal{I}_\sigma$  for  $\mathcal{I}(p_\sigma)$ . Now take  $h = \operatorname{Re} \mathcal{I}$  as a Morse function, and write  $h_\sigma$  for  $h(p_\sigma)$ . The critical points are all of the same index  $n$ , as explained in section 3.1.1. So  $h$  is a perfect Morse function, and can be used to describe the appropriate relative homology group in which the possible integration cycles take values. This group is  $H_n(X, X_{-T}; \mathbb{Z})$ , where  $X = \mathbb{C}^n$ , and  $X_{-T}$  is defined by the condition  $h \leq -T$ , for large  $T$ .

Introducing on  $\mathbb{C}^n$  the Kahler metric  $ds^2 = \sum_i |dx^i|^2$ , the downward flow equation (3.8) can be written as in (3.10):

$$\frac{dx^i}{dt} = -\frac{\partial \mathcal{I}}{\partial \bar{x}^i}. \quad (3.32)$$

For generic  $\lambda$ , there are no flows between critical points. This follows from the fact that  $\operatorname{Im} \mathcal{I}$  is conserved along a critical point, as shown in eqn. (3.11). (That computation generalizes immediately to  $n$  dimensions, as long as the metric on  $X$  is hermitian. Just as in the one-dimensional case, if  $X$  is Kahler, and we use the Kahler form of  $X$  as a symplectic form, then the gradient flow with respect to the Morse function  $\operatorname{Re} \mathcal{I}$  is the same as the Hamiltonian flow with Hamiltonian  $\operatorname{Im} \mathcal{I}$ ; naturally,  $\operatorname{Im} \mathcal{I}$  is conserved in this flow.) If  $p_\sigma$  and  $p_\tau$  are critical points with distinct values of  $\mathcal{I}$ , then for generic  $\lambda$  they have distinct values of  $\operatorname{Im} \mathcal{I}$  and there is no flow between them. But if  $\mathcal{I}_\sigma = \mathcal{I}_\tau$ , there is never a flow between  $p_\sigma$  and  $p_\tau$ , since  $h = \operatorname{Re} \mathcal{I}$  is always strictly decreasing along a non-trivial flow, as explained in eqn. (3.9).

We define a Stokes ray to be a ray along which  $\operatorname{Im} \mathcal{I}_\sigma = \operatorname{Im} \mathcal{I}_\tau$  for some  $\sigma, \tau \in \Sigma$ . There are only finitely many Stokes rays (we have assumed that  $\Sigma$  is a finite set, but even without this assumption, for any polynomial  $f$ ,  $\Sigma$  has only finitely many components, and  $\mathcal{I}$ , which is constant along each component, has only finitely many critical values).

Just as in the one-variable case, we attach to each critical point  $p_\sigma$  a cycle  $\mathcal{J}_\sigma$ , known as a Lefschetz thimble, by taking the union of all downward flows that start at  $p_\sigma$ , or equivalently the values at  $t = 0$  of downward flows  $u^i(t)$  that start at  $p_\sigma$  at  $t = -\infty$ . As long as  $\lambda$  is not on a Stokes ray,  $\mathcal{J}_\sigma$  is closed and has its ends at  $h = -\infty$ . Hence,  $\mathcal{J}_\sigma$  defines an element of the relative homology  $H_n(X, X_{-T}; \mathbb{Z})$ . Moreover, according to Morse theory, the  $\mathcal{J}_\sigma$  give a basis of this relative homology.

In crossing a Stokes ray along which  $\operatorname{Im} \mathcal{I}_\sigma = \operatorname{Im} \mathcal{I}_\tau$ , the cycles  $\mathcal{J}_\sigma$  and  $\mathcal{J}_\tau$  may jump. Such jumping involves nontrivial flows between  $p_\sigma$  and  $p_\tau$ . A downward flow from  $p_\sigma$  to  $p_\tau$  is only possible if  $h_\sigma > h_\tau$ . The general form of the jumping across a Stokes ray with

$\text{Im } \mathcal{I}_\sigma = \text{Im } \mathcal{I}_\tau$ ,  $h_\sigma > h_\tau$  is

$$\begin{aligned}\mathcal{J}_\sigma &\rightarrow \mathcal{J}_\sigma + \mathbf{m}_{\sigma\tau} \mathcal{J}_\tau \\ \mathcal{J}_\tau &\rightarrow \mathcal{J}_\tau.\end{aligned}\tag{3.33}$$

Here  $\mathbf{m}_{\sigma\tau}$  receives a contribution of  $\pm 1$  from each downward flow line from  $p_\sigma$  to  $p_\tau$ . (The sign of the contribution is the sign of a certain determinant, described at the end of section 4.1.2.)

Whenever the Lefschetz thimble  $\mathcal{J}_\sigma$  is defined, the integral

$$Q_{\sigma,\lambda} = (-i\lambda)^{n/2} \int_{\mathcal{J}_\sigma} d^n x \exp(\mathcal{I})\tag{3.34}$$

is always convergent. Moreover, the phase of the function  $\exp(\mathcal{I})$  is constant along the integration cycle, and its modulus has a unique maximum at the critical point  $p_\sigma$ . Hence, for large  $\lambda$ , this integral has an asymptotic expansion

$$Q_{\sigma,\lambda} \sim \exp(\mathcal{I}_\sigma) \sum_{t=0}^{\infty} a_t \lambda^{-t}.\tag{3.35}$$

To find the asymptotic behavior of the analytically continued integral, we will use the expansion

$$\mathcal{C} = \sum_{\sigma} \mathbf{n}_\sigma \mathcal{J}_\sigma.\tag{3.36}$$

For real  $\lambda$ ,  $\mathcal{C}$  is simply the original real integration cycle  $\mathcal{C}_{\mathbb{R}}$ . Unlike the case of the Airy function, we do not necessarily have a convenient way to directly determine the dependence of  $\mathcal{C}$  on  $\lambda$ . However, once the coefficients  $\mathbf{n}_\sigma$  are known for real  $\lambda$ , we can determine their dependence on  $\lambda$  by using the “wall-crossing” formulas (3.33).

To determine the coefficients for real  $\lambda$ , we proceed as in the one-variable case. For each critical point  $p_\sigma$ , we define a cycle  $\mathcal{K}_\sigma$  that consists of upward-going flow lines that start at  $p_\sigma$ . Away from Stokes rays,  $\mathcal{K}_\sigma$  is an element of  $H_n(X, X^T)$  for very large  $T$ , where  $X^T \subset X$  is defined by the condition  $h \geq T$ . Just as for  $n = 1$ , the groups  $H_n(X, X_{-T})$  and  $H_n(X, X^T)$  are naturally dual, and moreover the pairings of the  $\mathcal{J}_\sigma$  and  $\mathcal{K}_\tau$  are

$$\langle \mathcal{J}_\sigma, \mathcal{K}_\tau \rangle = \delta_{\sigma\tau}.\tag{3.37}$$

As a result, the coefficients in (3.36) are

$$\mathbf{n}_\sigma = \langle \mathcal{C}, \mathcal{K}_\sigma \rangle.\tag{3.38}$$

Thus,  $\mathbf{n}_\sigma$  can be computed by counting (with signs) upward flows from  $p_\sigma$  to  $\mathcal{C}$ .

Let us analyze this formula for real  $\lambda$  and the original integration cycle  $\mathcal{C}_{\mathbb{R}}$ . We divide the set  $\Sigma$  of critical points into three subsets:  $\Sigma_{\mathbb{R}}$  consists of critical points with all  $x^i$  real,  $\Sigma_-$  consists of critical points with  $x^i$  not all real and  $h < 0$ , and  $\Sigma_{\geq 0}$  consists of critical points with  $x^i$  not real and  $h \geq 0$ . Note that  $h = 0$  along  $\mathcal{C}_{\mathbb{R}}$ , and in particular  $h_{\sigma} = 0$  for  $\sigma \in \Sigma_{\mathbb{R}}$ . If  $\sigma \in \Sigma_{\mathbb{R}}$ , then the unique upward flow from  $p_{\sigma}$  to  $\mathcal{C}_{\mathbb{R}}$  is the trivial one, as  $h$  would increase along a nonconstant flow. So in this case  $\mathbf{n}_{\sigma} = 1$ . If  $\sigma \in \Sigma_{\geq 0}$ , there are no upward flow lines from  $p_{\sigma}$  to  $\mathcal{C}_{\mathbb{R}}$ , as this would contradict the fact that  $h$  increases along such lines. Finally, if  $\sigma \in \Sigma_-$ , there may or may not be ascending flow lines from  $p_{\sigma}$  to  $\mathcal{C}_{\mathbb{R}}$ .

The general structure of the expression for  $\mathcal{C}_{\mathbb{R}}$ , with real  $\lambda$ , is therefore

$$\mathcal{C}_{\mathbb{R}} = \sum_{\sigma \in \Sigma_{\mathbb{R}}} \mathcal{J}_{\sigma} + \sum_{\sigma \in \Sigma_-} \mathbf{n}_{\sigma} \mathcal{J}_{\sigma}, \quad (3.39)$$

with unknown integer coefficients  $\mathbf{n}_{\sigma}$ . Correspondingly, for real  $\lambda$ , our original integral  $Q_{\lambda}$  is

$$Q_{\lambda} = \sum_{\sigma \in \Sigma_{\mathbb{R}}} Q_{\sigma, \lambda} + \sum_{\sigma \in \Sigma_-} \mathbf{n}_{\sigma} Q_{\sigma, \lambda}. \quad (3.40)$$

Since each term on the right hand side has the asymptotic expansion (3.35), this formula makes manifest the behavior for large real  $\lambda$ . For complex  $\lambda$ , there is a similar formula, but of course one must use the appropriate coefficients.

The significance of the formula (3.40) is as follows. In studying the function  $Q_{\lambda}$  for large real  $\lambda$ , the obvious critical points that one would expect to contribute are the ones that lie on the real cycle  $\mathcal{C}_{\mathbb{R}}$ . They indeed all contribute with a coefficient of 1, as one would expect. Critical points in  $\Sigma_{\geq 0}$  would make contributions to the large  $\lambda$  behavior that typically would grow exponentially and dominate the real critical points (if  $h_{\sigma} > 0$ ) and that in the exceptional case  $h_{\sigma} = 0$  would compete for large real  $\lambda$  with the real critical points (which also have  $h = 0$ ). However, such critical points do not contribute to the integral. Finally, there may be contributions from complex critical points with  $h$  strictly negative. Their contributions are exponentially small for large real  $\lambda$ . But there is no easy recipe to determine which such critical points do contribute exponentially small corrections to  $Q_{\lambda}$ , and with what weights. Rather, for  $\sigma \in \Sigma_-$ , one simply has to determine the coefficient  $\mathbf{n}_{\sigma}$  by counting (with signs) the upward flow lines from  $p_{\sigma}$  to  $\mathcal{C}_{\mathbb{R}}$ .

### 3.2.1 Essential Singularities

Now we wish to address the following question: Under what conditions does the analytically continued integral  $Q_{\lambda}$  have an essential singularity at  $\lambda = \infty$ ?

There is one special case in which there is no such singularity. Suppose that, near infinity in the complex  $\lambda$  plane, the critical points  $p_{\sigma}$  for which the coefficients  $\mathbf{n}_{\sigma}$  are nonzero all

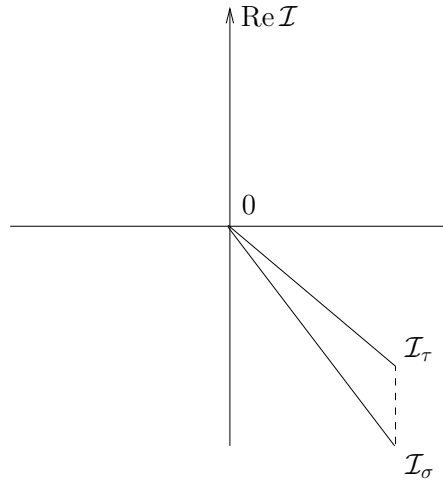


Figure 3: The complex  $\mathcal{I}$  plane is sketched here with  $h = \text{Re } \mathcal{I}$  running vertically. The exponents  $\mathcal{I}_\tau$  and  $\mathcal{I}_\sigma$  of two critical points  $p_\tau$  and  $p_\sigma$  have arguments between  $\pi$  and  $3\pi/2$ . If  $\text{Arg } \mathcal{I}_\tau > \text{Arg } \mathcal{I}_\sigma$ , then at a Stokes ray with  $\text{Im } \mathcal{I}_\tau = \text{Im } \mathcal{I}_\sigma$ , one has  $h_\tau > h_\sigma$ . This ensures that the coefficient  $\mathbf{n}_\tau$  is unaffected in crossing the Stokes line, though  $\mathbf{n}_\sigma$  may jump. The analytically continued integral will grow exponentially when  $\lambda$  is varied so that  $\text{Arg } \mathcal{I}_\tau$  exceeds  $3\pi/2$ .

have  $\mathcal{I}_\sigma = 0$ . Then  $Q_\lambda$  is bounded for  $\lambda \rightarrow \infty$ , in any direction. In particular, it has no essential singularity at infinity.

The hypothesis that the  $\mathcal{I}_\sigma = 0$  whenever  $\mathbf{n}_\sigma \neq 0$  is very strong, and there are not many obvious examples, other than an oscillatory Gaussian integral with  $f(x) = x^2$ . However, something similar is possible in Chern-Simons theory.

Now we want to explain a converse. For simplicity, assume that the values of  $\mathcal{I}_\sigma$  at critical points are distinct. (The purpose of this is to avoid possible cancellations.) Let  $p_\sigma$  be a critical point with  $\mathcal{I}_\sigma \neq 0$ , and suppose that in some wedge-shaped region of the complex  $\lambda$  plane,  $\mathbf{n}_\sigma \neq 0$ . We claim that in this case, the integral  $Q_\lambda$  has an essential singularity at  $\lambda = \infty$ . This result is immediate if  $h_\sigma > 0$ , for then the critical point  $p_\sigma$  makes an exponentially large contribution to  $Q_\lambda$  in the wedge-shaped region of the  $\lambda$  plane in which we have started.

Suppose, though, that in the initially considered region of the  $\lambda$  plane, all critical points with  $\mathbf{n}_\sigma \neq 0$  have  $h_\sigma \leq 0$ . This means that for such critical points,  $\mathcal{I}_\sigma = r_\sigma \exp(i\phi_\sigma)$ , with  $r_\sigma$  positive and  $\pi/2 \leq \phi_\sigma \leq 3\pi/2$ . Either the largest of the  $\phi_\sigma$  is at least  $\pi$  or the smallest is no greater than  $\pi$ ; suppose for definiteness we are in the former situation. (Otherwise, one makes the same argument, rotating  $\lambda$  in a counterclockwise rather than clockwise direction.) Let  $\tau$  be such that, in the initially chosen wedge in the complex  $\lambda$  plane,  $\phi_\tau$  is maximal among

critical points with nonzero  $\mathbf{n}_\tau$ . (If there is more than one critical point with maximal  $\phi$ , pick the one of smallest  $|\mathcal{I}|$ .) Increasing the argument of  $\lambda$  increases all  $\phi_\sigma$ 's by the same amount. Increase the argument of  $\lambda$  until  $\phi_\tau$  exceeds  $3\pi/2$ , so that  $h_\tau$  becomes positive. In the process, one may cross Stokes rays, but the fact that  $\phi_\tau$  is maximal and is at least  $\pi$  ensures that  $\mathbf{n}_\tau$  does not jump in crossing such rays. (See fig. 3; if  $\phi_\tau \geq \phi_\sigma$  and  $\text{Im } \mathcal{I}_\tau = \text{Im } \mathcal{I}_\sigma$ , then  $h_\tau \geq h_\sigma$  so  $\mathbf{n}_\tau$  does not jump.) Once  $h_\tau$  becomes positive,  $Q_\lambda$  grows exponentially with  $\lambda$ , ensuring that it has an essential singularity at  $\lambda = \infty$ .

### 3.2.2 A Special Situation

Now we will consider a special situation that was described in section 2.1. We assume that  $n = 2w$ , and that the polynomial  $f(x)$  is actually the imaginary part of a polynomial  $g$  in  $w$  complex variables  $z^1, \dots, z^w$ ; we suppose that  $g$  is generic enough to have only finitely many critical points, which are all nondegenerate. Then we consider the integral

$$Q_\lambda = \int d^w z d^w \bar{z} \exp(\lambda(g(z) - \overline{g(z)})), \quad (3.41)$$

initially for real  $\lambda$ .

Actually, we can generalize the problem slightly without any difficulty; the generalization is relevant to Chern-Simons theory. As  $\lambda$  is initially real, we can absorb it in the definition of  $g$ . Then instead of simply analytically continuing with respect to  $\lambda$ , we can analytically continue with respect to all of the coefficients in  $\bar{g}$ . In other words, the following integral

$$Q_{g,\tilde{g}} = \int_{\mathcal{C}_\mathbb{R}} d^w z d^w \tilde{z} \exp(g(z) - \tilde{g}(\tilde{z})), \quad (3.42)$$

is a convergent oscillatory integral if the polynomial  $\tilde{g}$  is the complex conjugate of the polynomial  $g$  (meaning that  $\tilde{g}$  is obtained from  $g$  by complex conjugating all coefficients) and the integration cycle  $\mathcal{C}_\mathbb{R}$  is defined by  $\tilde{z}^i = \bar{z}^i$ . We would like to analytically continue with respect to the coefficients in  $g$  and  $\tilde{g}$  to get a holomorphic function of those coefficients. (A similar generalization could be made in section 3.2; we could absorb the real number  $\lambda$  in the definition of the polynomial  $f$ , and then analytically continue with respect to all of the coefficients in  $f$ , not just  $\lambda$ .)

In our previous analysis, we had to avoid  $\lambda = 0$ , and similarly in the present context, we must impose a condition on how  $g$  and  $\tilde{g}$  are allowed to vary. The condition we want is that as  $g$  and  $\tilde{g}$  vary, the critical points of the polynomials  $g$  and  $\tilde{g}$  should remain nondegenerate and their number should remain fixed. For example, we do not allow leading coefficients in  $g$  or  $\tilde{g}$  to vanish in a way that would cause the number of critical points to become infinite or would cause some critical points to go to infinity. This restriction is assumed in what follows.



The basic idea behind the analytic continuation with respect to  $g$  and  $\tilde{g}$  is the same as we have already described. For  $\tilde{g} \neq \bar{g}$ , we will have to relax the condition  $\tilde{z} = \bar{z}$  and integrate over a more general cycle  $\mathcal{C}$ :

$$Q_{\mathcal{J};g,\tilde{g}} = \int_{\mathcal{C}} d^w z d^w \tilde{z} \exp(g(z) - \tilde{g}(\tilde{z})). \quad (3.43)$$

Of course, the integration cycle  $\mathcal{C}$  must be chosen so that the integral converges. By now we know how to describe cycles of the appropriate form. We set  $\mathcal{I} = g(z) - \tilde{g}(\tilde{z})$  and  $h = \text{Re } \mathcal{I}$ . We treat  $h$  as a Morse function. To each critical point  $p_\sigma$  of  $h$ , we attach a Lefschetz thimble  $\mathcal{J}_\sigma$ . These cycles form a basis for the appropriate relative homology group. In the original case  $\tilde{g} = \bar{g}$ , we express the original real integration cycle  $\mathcal{C}_{\mathbb{R}}$  as a linear combination of the  $\mathcal{J}_\sigma$ 's, by the usual formula  $\mathcal{C}_{\mathbb{R}} = \sum_{\sigma} \mathbf{n}_\sigma \mathcal{J}_\sigma$ . The integrals  $Q_{\mathcal{J}_\sigma;g,\tilde{g}}$  are convergent for all  $\sigma$ ,  $g$  and  $\tilde{g}$ , and the desired analytic continuation  $Q_{g,\tilde{g}}$  of our original integral is given by

$$Q_{g,\tilde{g}} = \sum_{\sigma} \mathbf{n}_\sigma Q_{\mathcal{J}_\sigma;g,\tilde{g}}. \quad (3.44)$$

The only subtlety is the usual one. As one varies  $g$  and  $\tilde{g}$  away from the locus  $\tilde{g} = \bar{g}$ , one may cross surfaces on which the cycles  $\mathcal{J}_\sigma$  and therefore the coefficients  $\mathbf{n}_\sigma$  jump. This happens when there are flows between different critical points, as a result of which the downward flow from the critical point with greater  $h$  fails to define a relative homology cycle. Since  $\text{Im } \mathcal{I}$  is a conserved quantity in the flow equation, the jumping can only occur when two distinct critical points have the same value of  $\text{Im } \mathcal{I}$ . The jumping takes the form  $\mathcal{J}_\sigma \rightarrow \mathcal{J}_\sigma + \mathbf{m}_{\sigma\tau} \mathcal{J}_\tau$ , where  $h_\sigma > h_\tau$  and  $\text{Im } \mathcal{I}_\sigma = \text{Im } \mathcal{I}_\tau$ . We will refer to surfaces on which distinct critical points have the same value of  $\text{Im } \mathcal{I}$  as Stokes surfaces. We also use the term Stokes curve for a Stokes surface of real dimension 1 (thus, a Stokes ray is a special case of a Stokes curve).

What is rather special about this problem is that  $\mathcal{I} = g(z) - \tilde{g}(\tilde{z})$  is the sum of a function on one space (which we will call  $\mathbb{C}^w$ , parametrized by the  $z^i$ ) and a function on another space (which we will call  $\tilde{\mathbb{C}}^w$ , a second copy of  $\mathbb{C}^w$  parametrized by the  $\tilde{z}^i$ ).  $\mathcal{I}$  is defined on the product space  $X = \mathbb{C}^{2w} = \mathbb{C}^w \times \tilde{\mathbb{C}}^w$ . The fact that  $\mathcal{I}$  is the sum of functions on the two factors has obvious implications for the critical points.

Let  $q_\sigma$ ,  $\sigma \in \Sigma_g$  be the critical points of  $g$ , as a function on  $\mathbb{C}^w$ , and let  $r_\tau$ ,  $\tau \in \Sigma_{\tilde{g}}$ , be the critical points of  $\tilde{g}$ , as a function on  $\tilde{\mathbb{C}}^w$ . Then the critical points of  $\mathcal{I}$  are the points  $p_{\sigma,\tau} = q_\sigma \times r_\tau \in \mathbb{C}^w \times \tilde{\mathbb{C}}^w = X$ . Similarly, the Lefschetz thimble  $\mathcal{J}_{\sigma,\tau}$  associated to a critical point  $p_{\sigma,\tau}$  in  $X$  is the product of cycles  $\mathcal{J}_\sigma^g \in \mathbb{C}^w$  and  $\mathcal{J}_\tau^{\tilde{g}} \in \tilde{\mathbb{C}}^w$ . Here  $\mathcal{J}_\sigma^g$  is the Lefschetz thimble in  $\mathbb{C}^w$  associated to the critical point  $q_\sigma$  of the Morse function  $\text{Re } g$ , and similarly  $\mathcal{J}_\tau^{\tilde{g}}$  is the Lefschetz thimble in  $\tilde{\mathbb{C}}^w$  associated to the critical point  $r_\tau$  of the Morse function  $\text{Re } (-\tilde{g})$ .

We can therefore rewrite (3.44) as follows:

$$Q_{g,\tilde{g}} = \sum_{\sigma \in \Sigma_g} \sum_{\tau \in \Sigma_{\tilde{g}}} n_{\sigma,\tau} \int_{\mathcal{J}_\sigma^g} d^w z \exp(g(z)) \int_{\mathcal{J}_\tau^{\tilde{g}}} d^w \tilde{z} \exp(-\tilde{g}(\tilde{z})). \quad (3.45)$$

Thus  $Q_{g,\tilde{g}}$  is everywhere a finite sum of functions of  $g$  times functions of  $\tilde{g}$ , obtained from integrals over the  $z$ 's and the  $\tilde{z}$ 's, respectively. The coefficients in the sum, however, can jump as one crosses Stokes surfaces.

As was explained in section 2, the problem that we have just described is a prototype for analytic continuation of Chern-Simons gauge theory with complex gauge group.

### 3.3 Symmetries

Since gauge symmetry is an essential ingredient in Chern-Simons theory, including symmetries in our finite-dimensional models will improve their analogy with Chern-Simons. Before taking this step, we consider a few preliminaries.

#### 3.3.1 Oscillatory Integrals On More General Manifolds

One generalization is to consider, instead of an oscillatory integral on  $\mathbb{R}^n$ , an oscillatory integral on some other space  $Y$ .  $Y$  must have a suitably nice complexification  $X$ , but we will not attempt to be precise about this. Instead we just describe a typical example. We take  $Y$  to be a real affine variety such as the one defined by the equation

$$x^2 + y^2 = z^2 + 1. \quad (3.46)$$

Then we introduce a sufficiently generic polynomial  $f(x, y, z)$ , and consider the oscillatory integral  $Z_\lambda = \int_Y d\theta \exp(i\lambda f)$ , where  $d\theta = dx dy/z$  is the natural volume form on  $Y$ .

Analytic continuation in  $\lambda$  can now be carried out by an obvious generalization of what we have said for  $\mathbb{R}^n$ . We embed  $Y$  in the complex manifold  $X$  defined by the same equation (3.46) but now with  $x, y$ , and  $z$  regarded as complex variables. The top degree form  $d\theta$  on  $Y$  extends to a holomorphic form of top degree on  $X$  (given by the same formula  $dx dy/z$ ), which can be integrated over any middle-dimensional cycle in  $X$ . Introducing the Morse function  $h = \text{Re}(i\lambda f)$ , where  $f$  is regarded now as a holomorphic function on  $X$ , we associate to each critical point a relative homology cycle obtained by downward flow. The analytic continuation is made as usual by expressing the original integration cycle  $Y$  in terms of these cycles, and then taking Stokes phenomena into account.

### 3.3.2 Manifolds Of Critical Points

Another generalization is to consider a polynomial  $f$  on  $\mathbb{R}^n$  (or on some more general space, as just described) that has not isolated critical points but a manifold  $\mathcal{N}$  of critical points. The analytic continuation of  $f$  to a holomorphic function on  $\mathbb{C}^n$  likewise has a manifold  $\mathcal{M}$  of critical points;  $\mathcal{M}$  is a complex manifold, of course, and its intersection with  $\mathbb{R}^n$  coincides with  $\mathcal{N}$ . (We will consider only the case that  $\mathcal{M}$  is smooth, and we assume that the matrix of second derivatives of  $f$  in the normal direction to  $\mathcal{M}$  is invertible.) We want to use Morse theory on  $\mathbb{C}^n$  to study the analytic continuation of the usual integral  $Z_\lambda = \int d^n x \exp(i\lambda f)$ .

Suppose that  $\mathcal{M}$  is a union of components  $\mathcal{M}_\alpha$  of complex dimension  $t_\alpha$ .  $\mathcal{I} = i\lambda f$  has a constant value  $\mathcal{I}_\alpha$  on  $\mathcal{M}_\alpha$ , since  $\mathcal{M}_\alpha$  is a critical point set. As usual, we consider  $h = \text{Re} \mathcal{I}$  as a Morse function. Then  $\mathcal{M}_\alpha$  has Morse index  $n - t_\alpha$ . (This is shown by essentially the same argument used in section 3.1.1 for the case  $t_\alpha = 0$ ; one can choose local coordinates in the normal direction to  $\mathcal{M}_\alpha$  so that  $\mathcal{I} = c + \sum_{i=1}^{n-t_\alpha} z_i^2$ .) The explanation in section 3.1.1 for why  $h$  is a perfect Morse function assumed that  $t_\alpha = 0$  for all components, so that all critical points have the same Morse index. Here is an alternative argument that does not assume  $t_\alpha = 0$ . In general, a Morse function is perfect if there are no flows between distinct critical points. (This is only possible for Morse functions that are unbounded below, like the ones we consider here, since otherwise any downward flow from one critical point ends at another critical point.) As we have found in eqn. (3.11),  $\text{Im} \mathcal{I}$  is a conserved quantity for the flow equation. Moreover, there never are flows between different critical points with the same value of  $\mathcal{I}$ . For generic  $\lambda$ , no two distinct components of  $\mathcal{M}_\alpha$  with different values of  $\mathcal{I}_\alpha$  have the same value of  $\text{Im} \mathcal{I}_\alpha$ , so generically there are no flows between distinct critical points and  $h$  is a perfect Morse function.

There is a general recipe, given a perfect Morse function on any manifold  $X$ , for describing the relative homology of  $X$  in terms of downwards flows from critical points. Let  $\mathcal{S}$  be a component of the critical point set that has Morse index  $s$ . Let  $\mathcal{W} \subset \mathcal{S}$  be a cycle of dimension  $r$ . For any point  $w \in \mathcal{W}$ , the points in  $X$  that lie on downwards flows from  $w$  form an  $s$ -dimensional subspace of  $X$ . So the points in  $X$  that can be reached by downward flow from somewhere on  $\mathcal{W}$  are a family of dimension  $r + s$ . This family is a cycle  $\mathcal{J}_\mathcal{W}$  of dimension  $r + s$  in the relative homology. These cycles, which we view as generalized Lefschetz thimbles, generate the relative homology of  $X$  (subject only to the relations that hold in the homology of the critical point set).

Applying this to our problem, if  $\mathcal{M}_\alpha$  is a component of the critical point set that has complex dimension  $t_\alpha$  and Morse index  $n - t_\alpha$ , then to get an integration cycle on  $X$  – which should be a cycle of dimension  $n$  – we need to start with a cycle on  $\mathcal{M}_\alpha$  of real dimension  $t_\alpha$ , that is, a middle-dimensional cycle.

So the analytic continuation of the integral  $Z_\lambda$  can be described using relative homology

cycles attached to the components  $\mathcal{M}_\alpha$  of the critical point set. But in general to a component  $\mathcal{M}_\alpha$  we attach not just one cycle but a number equal to the rank of the middle-dimensional homology of  $\mathcal{M}_\alpha$ .

This construction may have various applications to Chern-Simons theory, since in general the moduli space of representations of the fundamental group of a three-manifold has components of positive dimension. However, the only application we will make in this paper concerns reducible critical points, that is, flat connections that preserve some gauge symmetry. Moreover, this application concerns a situation in which every relevant component of  $\mathcal{M}_\alpha$  has middle-dimensional cohomology of rank 1, and hence contributes only one cycle to the middle-dimensional cohomology. In a sense, therefore, our application involves a situation that is as simple as the case of only isolated critical points.

### 3.3.3 Group Actions

Why might critical points not be isolated? One simple answer to this question will be important in our study of gauge theory. In the presence of a group action, the critical points will form orbits of the group, and this may force them to be non-isolated.

We consider a manifold  $Y$  with a chosen action of a compact connected Lie group  $H$ . We also assume that  $Y$  has a good complexification  $X$ , which possesses a real involution that leaves  $Y$  fixed. We endow  $Y$  with an  $H$ -invariant function  $f$  whose critical points consist of finitely many  $H$  orbits  $\mathcal{O}_\sigma$ ,  $\sigma \in \Sigma$ .

As usual, we want to study the analytic continuation of the integral  $Z_\lambda = \int_Y d\theta \exp(i\lambda f)$ , where  $d\theta$  is the restriction to  $Y$  of a holomorphic top form on  $X$ . Assuming that  $f$  can be analytically continued to a holomorphic function on  $Y$  (which we also denote as  $f$ ), we analytically continue the integral  $Z_\lambda$  by embedding  $Y$  in  $X$  and using the Morse function  $h = \operatorname{Re}(i\lambda f)$  on  $X$ . We assume that the  $H$  action on  $Y$  analytically continues to an action on  $X$  of a complex Lie group  $G$  that is the complexification of  $H$ ; in this case, the critical points of  $h$  are orbits of  $G$ . We assume that there are only finitely many such critical orbits, near each of which  $h$  is nondegenerate in the normal directions. For clarity, we will write  $\mathcal{O}^G$  or  $\mathcal{O}^H$  for, respectively, a critical orbit of  $G$  acting on  $X$  or a critical orbit of  $H$  acting on  $Y$ .

To orient ourselves, let us first consider the case that  $H$  acts freely on  $Y$ , and  $G$  acts freely on  $X$ . Then we can write  $Z_f$  as an integral over  $Y' = Y/H$ :

$$Z_f = \int_{Y'} d\theta' \exp(i\lambda f), \quad (3.47)$$

where the volume form  $d\theta'$  is obtained by integrating  $d\theta$  over the fibers of  $Y \rightarrow Y'$ . From

a “downstairs” point of view, to analytically continue  $Z_f$ , we must embed  $Y'$  in a suitable complex manifold  $X'$ . Roughly speaking, the appropriate choice is  $X' = X/G$ , which – as  $G$  is a complexification of  $H$  and  $X$  is a complexification of  $Y$  – is a complexification of  $Y' = Y/H$ . (Unstable orbits – the notion is discussed below – should be discarded in defining the quotient  $X/G$ .)

Since  $f$  has only finitely many critical orbits as a function on  $X$ , it follows that as a function on  $X'$ ,  $f$  has only finitely many critical points. We are thus in the familiar situation of isolated critical points only, and the relative homology of  $X'$  therefore has a basis consisting of a Lefschetz thimble  $\mathcal{J}'_\sigma$  for every critical point  $p'_\sigma$ . We can analytically continue the integral  $Z_f$  in the usual way by expressing the real cycle  $\mathcal{C}_\mathbb{R} = Y' \subset X'$  as a linear combination of the  $\mathcal{J}'_\sigma$ , and keeping track of Stokes phenomena.

How would we get an equivalent result by computing upstairs on  $Y$  and  $X$  rather than downstairs on  $Y'$  and  $X'$ ? Every critical point  $p'_\sigma$  on  $X'$  corresponds to a critical orbit  $\mathcal{O}^X_\sigma$  on  $X$ . Since we have assumed that  $G$  acts freely on  $X$ , the orbits are all copies of  $G$ . Topologically, the complex Lie group  $G$  is isomorphic to the cotangent bundle  $T^*H$  of its maximal compact subgroup  $H$ . The middle-dimensional homology of  $T^*H$  is of rank 1, generated by the zero section of the cotangent bundle. Hence, from an upstairs point of view, the middle-dimensional relative homology of  $X$  has a basis with one generator for every critical orbit. These generators correspond in an obvious way to the generators of the relative homology of  $X'$ . The associated Lefschetz thimbles also have an obvious correspondence, and the Stokes surfaces are the same whether one computes upstairs or downstairs. So we can perform more or less equivalent computations upstairs or downstairs.

In the downstairs description, the critical points are isolated, so we do not really need the theory sketched in section 3.3.2 as long as  $H$  acts freely. Now let us consider a more general case in which  $H$  does *not* act freely on  $Y$ . This being so, the quotient  $Y' = Y/H$  is singular, and the general theory would not readily apply to it. So instead we will calculate on  $Y$ . We still assume that the critical point set of  $f$  on  $Y$  is a finite union of  $H$  orbits  $\mathcal{O}^H_\sigma$ , and that the critical point set of  $f$  on  $X$  is an analogous (but perhaps larger<sup>12</sup>) finite union of  $G$  orbits  $\mathcal{O}^G_\tau$ .

Any critical orbit  $\mathcal{O}^H_\sigma$  in  $Y$  has a complexification that is one of the  $\mathcal{O}^G_\tau$ . The complexification of an  $H$  orbit  $\mathcal{O}^H \subset Y$  is, from a topological point of view, simply  $T^*\mathcal{O}^H$ . Its middle-dimensional homology is of rank 1, generated by the zero section of the cotangent bundle. Each critical orbit in  $Y$  therefore contributes one generator of the middle-dimensional relative homology of  $X$ , just as in the case of a free action.

What about critical orbits in  $X$  that do not intersect  $Y$ ? We will show that not all critical

---

<sup>12</sup>This is possible even in the absence of a group action. For example, the function  $f(x) = x^3/3 + x$  has complex critical points that do not lie on the real axis.

orbits need to be taken into account, but only certain “semistable” ones. The semistable orbits are all topologically of the form  $T^*\mathcal{O}^H$  for some  $H$  orbit  $\mathcal{O}^H$ , and again they have middle-dimensional homology of rank 1. (Every critical orbit that intersects  $Y$  is semistable.) So in short, the analytic continuation of  $Z_f$  is made using a middle-dimensional relative homology of  $X$  that has one generator for every semistable critical  $G$  orbit.

Now we will explain how semistability comes in. First of all, to construct the downward flowing cycles associated to a critical point, we need to pick a metric on  $X$ . We generally cannot pick this metric to be  $G$ -invariant, but it is commonly possible to pick an  $H$ -invariant Kahler metric, such that the Kahler form  $\omega$  is odd under the real involution of  $X$  that leaves  $Y$  fixed. (For example, if  $Y = \mathbb{R}^n$  with  $H$ -invariant metric  $\sum_i (dx^i)^2$ , then we take on  $X = \mathbb{C}^n$  the  $H$ -invariant Kahler metric  $\sum_i |dx^i|^2$ . In this case,  $\omega = -i \sum_k dx^k \wedge d\bar{x}^k$ , and is odd under  $x^k \leftrightarrow \bar{x}^k$ .) We assume the existence of such a metric.  $H$  acts on  $X$  preserving the symplectic form  $\omega$ , and it is possible to define a moment map  $\vec{\mu}$  for the action of  $H$  on  $X$ . Here  $\vec{\mu}$  maps a vector field  $V$  that generates a one-parameter subgroup of  $H$  to a function  $\mu_V$  that obeys

$$d\mu_V = \iota_V \omega, \quad (3.48)$$

where  $\iota_V$  is the operation of contraction with  $V$ . In indices

$$\frac{\partial \mu_V}{\partial x^i} = V^{\bar{j}} \omega_{\bar{j}i}, \quad \frac{\partial \mu}{\partial x^i} = V^j \omega_{j\bar{i}}. \quad (3.49)$$

(The map from  $V$  to  $\mu_V$  is a linear map from the Lie algebra of  $H$  to the space of real-valued smooth functions on  $X$ .) We fix possible additive constants in the  $\mu_V$  by requiring that these functions vanish identically on  $Y$ . This also ensures that the  $\mu_V$  are odd under complex conjugation. This follows from (3.48), since the vector fields  $V$  are even under complex conjugation and  $\omega$  is odd.

The downward flow equations of Morse theory have the beautiful property that  $\vec{\mu}$  is conserved, in other words  $\mu_V$  is conserved for all  $V$ . This follows from a straightforward computation. We have

$$\frac{d\mu_V}{dt} = \frac{\partial \mu_V}{\partial x^i} \frac{dx^i}{dt} + \frac{\partial \mu_V}{\partial \bar{x}^i} \frac{d\bar{x}^i}{dt} = -V^{\bar{j}} \omega_{\bar{j}i} g^{i\bar{k}} \frac{\partial h}{\partial x^k} - V^j \omega_{j\bar{i}} g^{\bar{i}k} \frac{\partial h}{\partial x^k}. \quad (3.50)$$

Since  $h$  is the real part of a holomorphic function  $\mathcal{I}$ , we have  $\partial h / \partial x^i = i \partial \text{Im } \mathcal{I} / \partial x^i$ ,  $\partial h / \partial \bar{x}^i = -i \partial \text{Im } \mathcal{I} / \partial \bar{x}^i$ . On the other hand, the Kahler form  $\omega$  and Kahler metric  $g$  obey  $\omega_{\bar{j}i} g^{i\bar{k}} = i \delta_{\bar{j}}^{\bar{k}}$ ,  $\omega_{j\bar{i}} g^{\bar{i}k} = -i \delta_j^k$ . So

$$\frac{d\mu_V}{dt} = -V^{\bar{j}} \frac{\partial \text{Im } \mathcal{I}}{\partial \bar{x}^{\bar{j}}} - V^j \frac{\partial \text{Im } \mathcal{I}}{\partial x^j} = -\iota_V d \text{Im } \mathcal{I}. \quad (3.51)$$

The right hand side vanishes, since  $\mathcal{I}$  is  $H$ -invariant.

A less computational explanation of this result can be found by observing that the gradient flow with respect to the Morse function  $\operatorname{Re}\mathcal{I}$  coincides with the Hamiltonian flow with Hamiltonian  $\operatorname{Im}\mathcal{I}$ . (Here we regard  $X$  as a symplectic manifold with symplectic form  $\omega$ .) As usual, the moment map for  $H$  is a conserved quantity for the flow generated by an  $H$ -invariant Hamiltonian.

We will call a critical  $G$ -orbit semistable if it contains at least one point with  $\vec{\mu} = 0$  and unstable otherwise. (If a semistable orbit is free or has only a finite stabilizer, we call it stable.) Since  $\vec{\mu}$  is conserved along flows, and vanishes identically on  $Y$ , any critical  $G$ -orbit that is connected to  $Y$  by a flow is semistable. So when we express the real integration cycle  $\mathcal{C}_{\mathbb{R}} = Y$  in terms of generalized Lefschetz thimbles associated to critical orbits, only semistable critical orbits enter. Similarly, as we explain momentarily, in crossing a Stokes surface, a cycle in the relative homology of  $X$  associated to a semistable critical orbit is always re-expressed in terms of other cycles associated to semistable critical orbits. So the analytic continuation of the integral  $Z_f$  can always be described using the semistable critical orbits only.

It remains to explain why Stokes phenomena connect semistable critical orbits only to each other. Suppose that  $\mathcal{O}^G$  is a semistable critical  $G$ -orbit. It contains a distinguished  $H$  orbit, namely the orbit  $\mathcal{W} = \mathcal{O}^H$  with  $\vec{\mu} = 0$ .  $\mathcal{O}^G$  is the complexification of  $\mathcal{O}^H$  and is topologically the cotangent bundle  $T^*\mathcal{O}^H$ . The middle-dimensional homology of  $\mathcal{O}^H$  is generated by  $\mathcal{W}$ . Since  $\vec{\mu}$  is conserved along flows, any critical orbit to which  $\mathcal{W}$  can be connected by a flow has a point with  $\vec{\mu} = 0$  and therefore is semistable.

In contrast to the middle-dimensional cohomology of a semistable  $G$  orbit, which is always of rank 1, an unstable  $G$  orbit might have vanishing or more complicated middle-dimensional cohomology. For  $G = SL(2, \mathbb{C})$ , an example of an unstable orbit is  $\mathbb{CP}^1$ , whose middle-dimensional cohomology vanishes. (Since  $\mathbb{CP}^1$  is a homogeneous space for  $H = SU(2)$ , and the condition  $\vec{\mu} = 0$  is  $H$ -invariant,  $\vec{\mu}$  would have to vanish everywhere on  $\mathbb{CP}^1$  if it vanishes anywhere; but if  $\vec{\mu}$  is identically zero, then (3.48) implies that  $V = 0$  for all generators of  $H$ , and the  $H$  action is trivial.) More generally, for any  $G$ , the flag manifold, which in general can have a middle-dimensional cohomology of high rank, is an example of an unstable orbit.

### 3.3.4 More on Group Actions

Continuing the discussion of the last subsection, let  $\mathcal{J}_{\sigma}$  be the Lefschetz thimble associated to a semistable critical orbit  $\mathcal{O}_{\sigma}^G$ .  $\mathcal{J}_{\sigma}$  consists of the points that can be reached by downward flow from  $\mathcal{O}_{\sigma}^H$ , the points in  $\mathcal{O}_{\sigma}^G$  with  $\vec{\mu} = 0$ .

Suppose that we are given an integration cycle  $\mathcal{C}$  and we wish to express it in terms of

the  $\mathcal{J}_\sigma$ .  $\mathcal{C}$  is a middle-dimensional,  $H$ -invariant element of the relative homology, and in practice one is usually interested in the case that  $\mathcal{C}$  lies in  $\vec{\mu}^{-1}(0)$ .

To find the coefficients in an expansion

$$\mathcal{C} = \sum_{\sigma} \mathbf{n}_{\sigma} \mathcal{J}_{\sigma}, \quad (3.52)$$

one wants to intersect  $\mathcal{C}$  with a dual set of cycles  $\mathcal{K}_{\sigma}$ . The dual cycles should have a natural pairing with the  $\mathcal{J}_{\sigma}$ , such that

$$\langle \mathcal{J}_{\sigma}, \mathcal{K}_{\tau} \rangle = \delta_{\sigma\tau}. \quad (3.53)$$

As we have defined  $\mathcal{J}_{\sigma}$  by downward flow from  $\mathcal{O}_{\sigma}^H \subset \mathcal{O}_{\sigma}^G$ , one's first thought is to define  $\mathcal{K}_{\sigma}$  by upward flow from  $\mathcal{O}_{\sigma}^H$ . This does not work well, since in this case the intersection of  $\mathcal{J}_{\sigma}$  with  $\mathcal{K}_{\sigma}$ , rather than a single point, would be the orbit  $\mathcal{O}_{\sigma}^H$ . (The algebraic intersection number would equal the Euler characteristic of this orbit; for instance, it would vanish for the case of a free orbit.) What one must do instead is to define  $\mathcal{K}_{\sigma}$  by upward flow starting from a suitable dual to  $\mathcal{O}_{\sigma}^H \subset \mathcal{O}_{\sigma}^G$ . The appropriate dual is a fiber of the cotangent bundle. In other words, identifying  $\mathcal{O}_{\sigma}^G$  as  $T^*\mathcal{O}_{\sigma}^H$ , we pick an arbitrary point  $q_{\sigma} \in \mathcal{O}_{\sigma}^H$  and write  $\tilde{\mathcal{O}}_{\sigma}^H$  for the fiber of  $T^*\mathcal{O}_{\sigma}^H$  at  $q_{\sigma}$ . Then  $\mathcal{K}_{\sigma}$ , defined by upward flow from  $\tilde{\mathcal{O}}_{\sigma}^H$ , meets  $\mathcal{J}_{\sigma}$  precisely at  $q_{\sigma}$ . ( $\mathcal{K}_{\sigma}$  lies in a dual version of the relative homology, spanned by cycles that are allowed to go to infinity along  $G$  orbits as well as in the region where the Morse function becomes large.) Moreover, away from a Stokes curve,  $\mathcal{K}_{\sigma}$  meets  $\mathcal{J}_{\tau}$  not at all if  $\sigma \neq \tau$ . So  $\langle \mathcal{J}_{\sigma}, \mathcal{K}_{\tau} \rangle = \delta_{\sigma\tau}$ , and the coefficients in (3.52) are  $\mathbf{n}_{\sigma} = \langle \mathcal{C}, \mathcal{K}_{\sigma} \rangle$ .

Similar questions arise when one considers the jumping phenomena  $\mathcal{J}_{\sigma} \rightarrow \mathcal{J}_{\sigma} + \sum_{\tau} \mathbf{m}_{\sigma\tau} \mathcal{J}_{\tau}$  in crossing a Stokes curve. The coefficients  $\mathbf{m}_{\sigma\tau}$  is computed by counting downward flow lines from  $\mathcal{O}_{\sigma}^H$  to  $\tilde{\mathcal{O}}_{\tau}^H$ . (Each such flow line is weighted by the sign of a certain determinant, as described in section 4.1.2.) The asymmetry between the definitions of downwards and upwards flowing cycles has an interesting consequence, which we will describe for the special case that there are two types of critical orbits – trivial orbits and free orbits. Suppose that the critical point set of an  $H$ -invariant Morse function  $h = \text{Re} \mathcal{I}$  contains a critical point  $p$  (located at  $\vec{\mu} = 0$ ) and a free stable  $G$  orbit  $\mathcal{O}^G$ . In general (when allowed by conservation of  $\text{Im } I$ ), there may be flows from  $p$  to  $\tilde{\mathcal{O}}^H$ , and the coefficient  $\mathbf{m}_{p\mathcal{O}}$  that appears in a jumping formula  $\mathcal{J}_p \rightarrow \mathcal{J}_p + \mathbf{m}_{p\mathcal{O}} \mathcal{J}_{\mathcal{O}}$  may be nonzero. However, the coefficient  $\mathbf{m}_{\mathcal{O}p}$  that counts flows from  $\mathcal{O}^H$  to  $p$  always vanishes. A quick way to show this is to observe that by replacing  $\mathcal{O}^H$  by an everywhere nonzero section of the cotangent bundle  $\mathcal{O}^G = T^*\mathcal{O}^H \rightarrow \mathcal{O}^H$ , such as the section  $\vec{\mu}^{-1}(c)$ , where  $c$  is a nonzero constant, one can eliminate all flows from  $\mathcal{O}^H$  to  $p$ . So the algebraic invariant counting such flows must vanish.

In a more general case in which the critical point set consists of several isolated points  $p_{\sigma}$  and several free orbits  $\mathcal{O}_{\alpha}$ , the coefficients  $\mathbf{m}_{\mathcal{O}_{\alpha}, p_{\sigma}}$  will vanish by the argument just indicated. (All other coefficients may be nonzero in general.) This means that the subspace of the relative homology generated by the cycles  $\mathcal{J}_{\mathcal{O}_{\alpha}}$  is invariant under all jumping phenomena.



The subspace generated by the cycles  $\mathcal{J}_{p_\sigma}$  is not invariant under jumping. This perhaps surprising fact will be demonstrated in an explicit and useful example in section 3.4.2.

### 3.4 Singularities

In studying an oscillatory integral  $\int d^n x \exp(i\lambda f)$ , for a single and fairly generic polynomial  $f$ , it is reasonable to expect that the critical points of  $f$  are isolated and irreducible, as we have assumed so far. But as soon as  $f$  depends on one or more parameters, one should expect to meet singularities, where two or more critical points meet.

Singularities play an important role because Stokes surfaces always pass through the locus (in the parameter space of the polynomial  $f$ ) at which two or more critical points meet. If we vary a parameter in  $f$  so that two critical points  $p_\sigma$  and  $p_\tau$  coincide, then  $f(p_\sigma) = f(p_\tau)$ , which certainly ensures that  $\text{Im}(i\lambda f(p_\sigma)) = \text{Im}(i\lambda f(p_\tau))$ . So a Stokes surface always passes through the singular locus with  $p_\sigma = p_\tau$ .

Moreover, the flow equations simplify at a singularity. To understand what happens to the Lefschetz thimbles in crossing a Stokes surface, in general we must solve the flow equation in  $n$  complex dimensions. However, near a singularity, most of the  $n$  complex variables are unimportant. If a singularity can be modeled by the behavior of a complex polynomial in  $k$  variables – where in practice  $k$  will often be 1 or 2 – then near the singularity, the flow equations reduce effectively to equations for those  $k$  variables; the other modes are “massive” and decay quickly in solving the flow equations.

The relation between singularities and Stokes phenomena is therefore very useful in practice. Indeed, the subject reviewed in this section was developed partly for analyzing singularities and their relation to oscillatory integrals [35].

In section 5, where we study the analytic continuation of the colored Jones polynomial, we will vary one parameter (the ratio of the highest weight of a representation to the Chern-Simons level) and we will encounter three rather generic types of singularity. These will be described here.

#### 3.4.1 More on the Airy Function

The only singularity that arises generically in varying a single complex parameter in the absence of any symmetries is a meeting of two critical points. Even if this phenomenon happens in complex dimension  $n$ , only one complex variable is essential in describing it. We

simply consider a cubic polynomial that depends on a parameter  $\epsilon$ :

$$f(x) = \frac{x^3}{3} - \epsilon x. \quad (3.54)$$

There are two critical points at  $x = \pm\sqrt{\epsilon}$ ; they meet at  $\epsilon = 0$ .

The corresponding oscillatory integral  $I(\lambda, \epsilon) = \int dx \exp(i\lambda f)$  is essentially the Airy function. We can use the scaling  $x \rightarrow x\lambda^{-1/3}$  to express this integral in terms of a function of one variable:

$$I(\lambda, \epsilon) = \lambda^{-1/3} I(1, \lambda^{2/3} \epsilon) = \lambda^{-1/3} \int dx \exp(i(x^3 - \lambda^{2/3} \epsilon x)). \quad (3.55)$$

Setting  $\epsilon = 0$ , a special case is  $I(\lambda, 0) \sim \lambda^{-1/3}$ . Comparing this to a Gaussian integral  $\tilde{I}(\lambda) = \int dx \exp(i\lambda x^2) \sim \lambda^{-1/2}$ , the ratio is

$$\frac{I(\lambda, 0)}{\tilde{I}(\lambda)} \sim \lambda^{1/6}, \quad \lambda \rightarrow \infty. \quad (3.56)$$

These facts will be useful in section 5.3.2.

For our remaining remarks, we use the above scaling to reduce to the case  $\lambda = 1$ . The good regions at infinity in the complex  $x$ -plane where  $\exp(if(x))$  decays are then independent of  $\epsilon$  and were sketched in fig. 1. The Lefschetz thimble associated to a critical point always connects two of the good regions.

Since the values of  $\mathcal{I} = if$  at the critical points are  $\mp 2\epsilon^{3/2}/3$ , the condition for a Stokes curve is that  $\epsilon^{3/2}$  should be imaginary; thus  $\text{Arg } \epsilon$  should equal  $\pi$  or  $\pm\pi/3$ . In particular, several Stokes curves meet at  $\epsilon = 0$ . It is clear why this is so: at  $\epsilon = 0$ , the two critical points meet, and therefore have the same value of  $\mathcal{I}$  and in particular the same value of  $\text{Im } \mathcal{I}$ .

Our application will involve the case that  $\epsilon$  is real. If  $\epsilon$  is positive, we write  $p_{\pm}$  for the critical points at  $x = \pm\sqrt{\epsilon}$ , and  $\mathcal{J}_{\pm}$  for the corresponding Lefschetz thimbles. If  $\epsilon$  is negative, we write  $\tilde{p}_{\pm}$  for the critical points at which the action is positive or negative, respectively. Thus the points  $\tilde{p}_{\pm}$  are located at  $x = \mp i\sqrt{-\epsilon}$ , and the action takes the values  $\mathcal{I}_{\pm} = \pm(2/3)(-\epsilon)^{3/2}$  at those points. We write  $\tilde{\mathcal{J}}_{-}$  for the Lefschetz thimble associated to  $\tilde{p}_{-}$ . The Lefschetz thimble  $\tilde{\mathcal{J}}_{+}$  associated to  $\tilde{p}_{+}$  is ill-defined if  $\epsilon < 0$ , because the ray  $\epsilon < 0$  is a Stokes ray. We define  $\tilde{\mathcal{J}}_{+}$  by taking the limit as  $\epsilon$  approaches the negative axis from above.

It is not difficult to describe the cycles  $\mathcal{J}_{\pm}$  and  $\tilde{\mathcal{J}}_{\pm}$  qualitatively as curves in the complex  $x$ -plane (fig. 4). From the figure, we see that the relation between these two bases for the

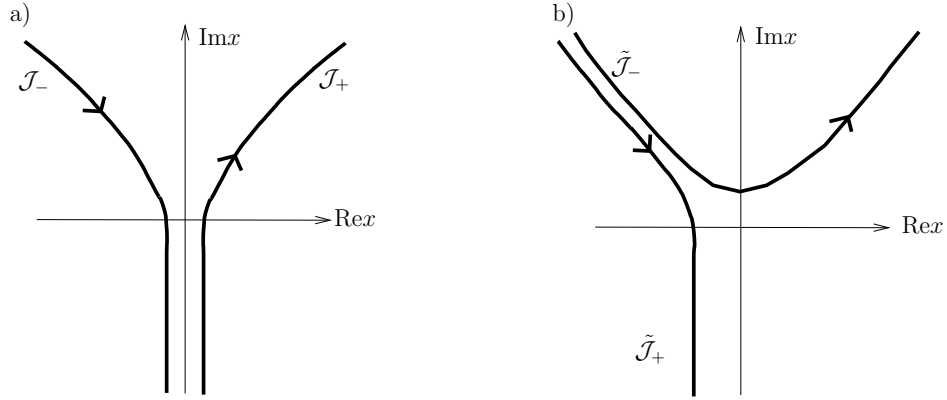


Figure 4: Qualitative behavior of Lefschetz thimbles for the Airy function for  $\epsilon > 0$  (a) and  $\epsilon < 0$  (b). In case (b),  $\epsilon$  has been given a small imaginary part, since the negative  $\epsilon$  axis is a Stokes curve.

relative homology is

$$\begin{aligned}\mathcal{J}_+ &= -\tilde{\mathcal{J}}_+ + \tilde{\mathcal{J}}_- \\ \mathcal{J}_- &= \tilde{\mathcal{J}}_+\end{aligned}\tag{3.57}$$

or equivalently

$$\begin{aligned}\tilde{\mathcal{J}}_+ &= \mathcal{J}_- \\ \tilde{\mathcal{J}}_- &= \mathcal{J}_+ + \mathcal{J}_-.\end{aligned}\tag{3.58}$$

### 3.4.2 An Example With $SO(2)$ Symmetry

Now we will consider a singularity that can arise by varying one parameter in the presence of  $SO(2)$  symmetry. A minimal case is a singularity that can be modeled by two variables  $x, y$ , with  $SO(2)$  acting on the  $xy$  plane by rotations. The basic invariant is  $x^2 + y^2$ . The origin  $x = y = 0$  is a critical point of any  $SO(2)$ -invariant polynomial. There may also be critical orbits characterized by nonzero values of  $x^2 + y^2$ . A meeting of two critical orbits both with  $x^2 + y^2 \neq 0$  can be modeled again by the Airy integral – the  $SO(2)$  action just factors out. The new case is a collision between a critical orbit with  $x^2 + y^2 \neq 0$  and the critical point at the origin.

Like a collision of two critical points in the absence of any symmetry, this can happen naturally as a result of adjusting one parameter. We simply consider the polynomial  $f(x, y) = (x^2 + y^2)^2/2 - \epsilon(x^2 + y^2)$ . There is a critical point at  $x = y = 0$  and a critical orbit  $x^2 + y^2 = \epsilon$ . We denote these as  $p$  and  $\mathcal{O}$  respectively.  $\mathcal{O}$  contains real points if  $\epsilon > 0$  but not if  $\epsilon < 0$ . As a real critical orbit,  $\mathcal{O}$  thus disappears in passing through  $\epsilon = 0$ . Setting  $\mathcal{I} = i\lambda f$  with

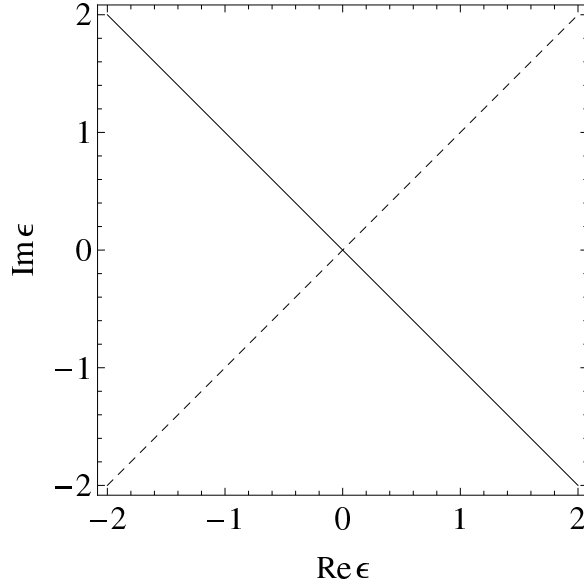


Figure 5: The two lines  $\epsilon = r \exp(\pm\pi i/2)$ , with real  $r$ , are Stokes lines in the sense that flows between the critical point  $p$  and the critical orbit  $\mathcal{O}$  are possible on these lines. Jumping, however, occurs only in crossing the line  $\epsilon = r \exp(-\pi i/2)$ , shown here as a solid line. There is no jumping of Lefschetz thimbles in crossing the dotted line.

$\lambda > 0$ , the values of  $\mathcal{I}$  on the critical orbits are  $\mathcal{I}_p = 0$  and  $\mathcal{I}_{\mathcal{O}} = -i\lambda\epsilon^2/2$ . So the Stokes curves are given by  $\text{Re } \epsilon^2 = 0$ , which corresponds to the pair of lines  $\epsilon = r \exp(\pm\pi i/2)$ , with real  $r$  (fig. 5). As expected, they meet at the singularity at  $\epsilon = 0$ .

A simple scaling reduces  $I(\lambda, \epsilon) = \int dx dy \exp\left(i\lambda((x^2 + y^2)^2 - \epsilon(x^2 + y^2))\right)$  to a function of one variable. Setting  $x, y \rightarrow \lambda^{-1/4}x, \lambda^{-1/4}y$ , we get

$$I(\lambda, \epsilon) = \lambda^{-1/2} I(1, \lambda^{1/2}\epsilon) = \lambda^{-1/2} \int dx dy \exp\left(i((x^2 + y^2)^2 - \lambda^{1/2}\epsilon(x^2 + y^2))\right). \quad (3.59)$$

Setting  $\epsilon = 0$ , a special case is  $I(\lambda, 0) \sim \lambda^{-1/2}$ . Let us compare this to the large  $\lambda$  behavior for  $\epsilon \neq 0$ . Assuming that (as in our application in eqn. (5.24)), the dominant contribution comes from the isolated, nondegenerate critical point at  $x = y = 0$ , a simple scaling shows that  $I(\lambda, \epsilon) \sim 1/\lambda$  for  $\epsilon \neq 0$ , so the analog of (3.56) is

$$\frac{I(\lambda, 0)}{I(\lambda, \epsilon)|_{\epsilon \neq 0}} \sim \lambda^{1/2}, \quad \lambda \rightarrow \infty. \quad (3.60)$$

By introducing polar coordinates in the  $xy$  plane,  $I(\lambda, \epsilon)$  can be expressed in terms of the error function.

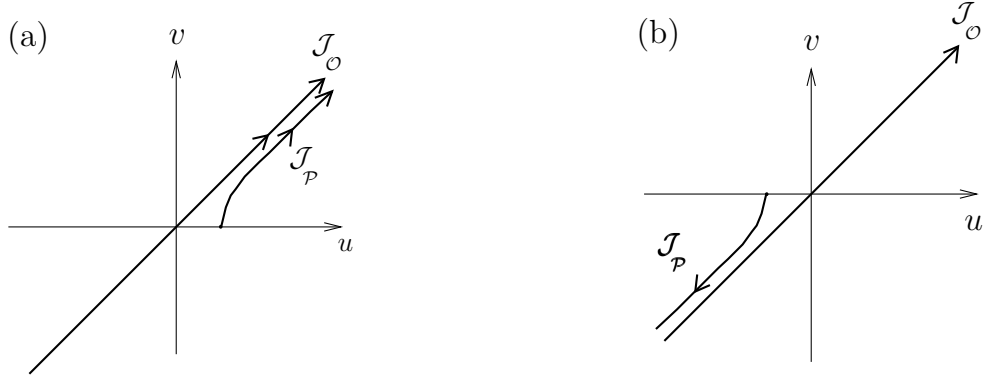


Figure 6: For real  $\epsilon$ , the cycle  $\mathcal{J}_p$  is derived from a curve in the  $uv$  plane that starts at  $(u, v) = (-\epsilon, 0)$  and consists of half of one branch of the hyperboloid  $u^2 - v^2 = \epsilon^2$ . We pick the half that is asymptotic to the curve  $u = v$ . For  $\epsilon < 0$ , the relevant half is asymptotic to  $u = v = \infty$ , as sketched in (a). For  $\epsilon > 0$ , it is asymptotic to  $u = v = -\infty$ , as sketched in (b). Accordingly, the cycle  $\mathcal{J}_p$  jumps when  $\epsilon$  changes sign. The jumping is by  $\mathcal{J}_p \rightarrow \mathcal{J}_p + \mathcal{J}_O$ , where  $\mathcal{J}_O$  is associated to the line  $u = v$  in the  $uv$  plane, irrespective of the sign of  $\epsilon$ . In other words,  $\mathcal{J}_p$  in (a) or  $\mathcal{J}_p + \mathcal{J}_O$  in (b) is a path from  $(u, v) = (-\epsilon, 0)$  to  $u = v = \infty$ .

The Lefschetz thimbles associated to critical orbits can be explicitly described in this example, because the symmetry can be used to reduce everything to complex dimension 1. Picking the Kahler metric  $ds^2 = |dx|^2 + |dy|^2$ , the moment map for the action of  $SO(2)$  becomes  $\mu = -i(x\bar{y} - y\bar{x})$ . Setting  $\mu = 0$ , we find that by an  $SO(2)$  rotation we can set  $y = 0$ . The Lefschetz thimbles that lie at  $\mu = 0$  are therefore surfaces of revolution that can be constructed by acting with the group  $SO(2)$  on a real curve in the complex  $x$  plane. For the rest of our analysis, we scale  $x$  to set  $\lambda = 1$ .

The flow equation has a conserved quantity  $\text{Im}(if)$ , so to find the right curve in the  $x$ -plane, we do not need to solve flow equations; it suffices to set  $\text{Im}(if)$  to a constant. Equivalently, we must set  $\text{Re}(x^2 - \epsilon)^2$  to a constant  $c$ . Setting  $x^2 = \epsilon + u + iv$ , the condition becomes  $u^2 - v^2 = c$ . For the Lefschetz thimble  $\mathcal{J}_O$  that flows down from the critical orbit  $x^2 + y^2 = \epsilon$ , we set  $c = 0$ , and choose the branch  $u = v$  to ensure that the real part of  $if(x, y)$  is negative. The Lefschetz thimble  $\mathcal{J}_p$  that flows down from the critical point  $x = y = 0$  passes through  $\epsilon + u + iv = 0$ , so the constant  $u^2 - v^2 = \text{Re}((u + iv)^2)$  is equal to  $\text{Re}(\epsilon^2)$ . The hyperboloid  $u^2 - v^2 = \text{Re}(\epsilon^2)$  has two branches. We pick the branch that passes through  $(u, v) = (-\epsilon, 0)$  (in other words,  $x = 0$ ), and then  $\mathcal{J}_p$  is derived from the half of this branch that is asymptotic to  $u = v$ . As shown in fig. 6, this branch is asymptotic to  $u, v \gg 0$  or  $u, v \ll 0$  depending on  $\epsilon$ . In particular, if  $\epsilon$  is real,  $\mathcal{J}_p$  has different asymptotic behavior depending on the sign of  $\epsilon$ .

From the figure, we can read off that between  $\epsilon < 0$  and  $\epsilon > 0$ , the Lefschetz thimbles change by

$$\begin{aligned}\mathcal{J}_p &\rightarrow \mathcal{J}_p + \mathcal{J}_\mathcal{O} \\ \mathcal{J}_\mathcal{O} &\rightarrow \mathcal{J}_\mathcal{O}.\end{aligned}\tag{3.61}$$

The meaning of the second statement is that the curve  $\mathcal{J}_\mathcal{O}$ , which is the diagonal  $u = v$ , is unchanged in going from fig. 6(a) to fig. 6(b). But  $\mathcal{J}_p$  in fig. 6(a) is a path from  $(u, v) = (-\epsilon, 0)$  to  $u = v = +\infty$ , while in fig. 6, it runs from  $(-\epsilon, 0)$  to  $u = v = -\infty$ . In fig. 6(b), to get a path from  $(-\epsilon, 0)$  to  $(+\infty, +\infty)$ , we need to take the sum  $\mathcal{J}_p + \mathcal{J}_\mathcal{O}$ . So  $\mathcal{J}_p$  jumps in passing through  $\epsilon = 0$ , though  $\mathcal{J}_\mathcal{O}$  does not. Invariance of  $\mathcal{J}_\mathcal{O}$  has a more general explanation that was described in section 3.3.4.

To recover the Lefschetz thimbles  $\mathcal{J}_p$  and  $\mathcal{J}_\mathcal{O}$ , we have to “rotate” (via  $SO(2)$ ) the curves that are sketched in fig. 6. In the case of  $\mathcal{J}_p$ , the relevant curve is topologically a half-line whose endpoint is  $SO(2)$ -invariant. Rotating it by  $SO(2)$ , we find that as a manifold with  $SO(2)$  action,  $\mathcal{J}_p$  is equivalent to  $\mathbb{R}^2$ ;  $SO(2)$  acts on  $\mathcal{J}_p$  with a fixed point at the origin that corresponds to the endpoint of the half-line. On the other hand, in the case of  $\mathcal{J}_\mathcal{O}$ , the relevant curve is a copy of  $\mathbb{R}$ , with its ends at infinity. Rotating it by  $SO(2)$ , we find that as a manifold with  $SO(2)$  action,  $\mathcal{J}_\mathcal{O}$  is equivalent to  $\mathbb{R} \times S^1$  with a free action of  $SO(2)$  on the second factor.

It is convenient to express (3.61) in a dual language. Given a general cycle  $\mathcal{C} = \mathbf{n}_p \mathcal{J}_p + \mathbf{n}_\mathcal{O} \mathcal{J}_\mathcal{O}$ , to compensate for the jumping (3.61) and ensure that  $\mathcal{C}$  varies smoothly, the coefficients jump by

$$\begin{aligned}\mathbf{n}_p &\rightarrow \mathbf{n}_p \\ \mathbf{n}_\mathcal{O} &\rightarrow -\mathbf{n}_p + \mathbf{n}_\mathcal{O}.\end{aligned}\tag{3.62}$$

Finally, let us return to the Stokes lines of fig. 5. We write  $l_+$  and  $l_-$  for the two Stokes lines given by  $\epsilon = r \exp(\pm i\pi/2)$  with real  $r$ . Scaling to  $\lambda = 1$ , the Morse function  $\text{Re}(if)$  vanishes at  $p$  and equals  $\pm r^2/2$  on  $l_\pm$ . So on  $l_+$ , there can be a downward flow from  $\mathcal{O}$  to  $p$ , and on  $l_-$  there can be a downward flow from  $p$  to  $\mathcal{O}$ . Flows of each type actually exist, as the reader can verify by slightly extending the analysis of fig. 6. ( $\mathcal{J}_\mathcal{O}$  always corresponds to the line  $u = v$ , and  $\mathcal{J}_p$  to part of a hyperboloid through  $u + iv = -\epsilon$ .) In general, the jumping in eqn. (3.61) occurs in crossing  $l_-$ , but there is no such jumping in crossing  $l_+$ . The algebraic invariant that “counts” the downward flows from  $\mathcal{O}$  to  $p$  along  $l_+$  vanishes (even though such flows exist), for a reason explained in section 3.3.4.

### 3.4.3 One More Case

The last singularity that we will encounter is more surprising. We will describe it first in a language with  $SO(1,1)$  rather than  $SO(2)$  symmetry. We consider  $SO(1,1)$  acting on real variables  $u, v$  by  $u \rightarrow ru, v \rightarrow r^{-1}v$ , with  $r$  real. The basic invariant is  $uv$ . We consider an integral  $\int du dv \exp(i\lambda f(uv))$  where  $f$  is a function of  $uv$ . The obvious case is that  $f$  is a quadratic function. However, and surprisingly, in the world of knots in  $S^3$ , there can be a natural reason for  $f$  to be an odd function of  $uv$ .

Under these conditions, the basic example of a function exhibiting critical behavior is

$$f(u, v) = (uv)^3/3 - \epsilon(uv). \quad (3.63)$$

Apart from the critical point  $p$  at  $u = v = 0$ , there are two critical orbits  $\mathcal{O}_\pm$  at  $uv = \pm\epsilon^{1/2}$ . For real  $\epsilon$ , these critical orbits are real or imaginary depending on the sign of  $\epsilon$ . (As real orbits, these orbits are hyperboloids with two topological components.)

A closely related problem is the case of  $SO(2)$  action on the  $xy$  plane with  $f(x, y) = (x^2 + y^2)^3/3 - \epsilon(x^2 + y^2)$ . The relation becomes apparent upon complexification; one problem can be mapped into the other by taking  $u = x + iy, v = x - iy$ . The problem of  $SO(1,1)$  acting on the real  $uv$  plane and the problem of  $SO(2)$  acting on the real  $xy$  plane correspond to two real slices of the same situation with complex variables. One real slice is defined by  $u = \bar{v}, v = \bar{u}$  and the other by  $u = \bar{v}$ . To study analytic continuation, as usual we must complexify the variables  $u$  and  $v$ , and then the  $SO(1,1)$  and  $SO(2)$  problems become equivalent. It is convenient to work with  $x$  and  $y$ . (We introduced  $u$  and  $v$  because the real slice with  $SO(1,1)$  symmetry will appear naturally in our application.)

First, we will determine the  $\lambda$  dependence of  $I(\lambda, \epsilon) = \int dx dy \exp(i\lambda f(x, y))$  using scaling symmetry. An obvious scaling of  $x$  and  $y$  shows that

$$I(\lambda, \epsilon) = \lambda^{-1/3} I(1, \lambda^{2/3} \epsilon) = \lambda^{-1/3} \int dx dy \exp(i((x^2 + y^2)^3/3 - \lambda^{2/3} \epsilon(x^2 + y^2))). \quad (3.64)$$

A special case is that  $I(\lambda, 0) \sim \lambda^{-1/3}$ . On the other hand, for  $\epsilon \neq 0$ , and choosing  $\epsilon < 0$  so that the only real critical point is the one at  $x = y = 0$ , a similar scaling gives  $I(\lambda, \epsilon) \sim \lambda^{-1}$ . So

$$\frac{I(\lambda, 0)}{I(\lambda, \epsilon)|_{\epsilon < 0}} \sim \lambda^{2/3}, \quad \lambda \rightarrow \infty. \quad (3.65)$$

By introducing polar coordinates in the  $xy$  plane, one can express  $I(\lambda, \epsilon)$  as a multiple of the Airy function. The main difference is that there is an extra possible integration cycle for  $I(\lambda, \epsilon)$ , as we now explain.

To study the analytic continuation systematically, we consider the integration cycles associated to critical orbits. Let  $\mathcal{J}_0, \mathcal{J}_+$ , and  $\mathcal{J}_-$  be the Lefschetz thimbles associated,

respectively, to the critical point at  $x = y = 0$  and to the critical orbits  $x^2 + y^2 = \pm\epsilon^{1/2}$ . The same reasoning as in section 3.4.2 can be used to describe them by applying an  $SO(2)$  rotation to suitable curves in the complex  $x^2$ -plane. In fact, apart from a substitution  $x \rightarrow x^2$ , the polynomial  $f(x, 0) = (x^2)^3/3 - \epsilon x^2$  is familiar from our study of the Airy function. Accordingly, the Lefschetz thimbles  $\mathcal{J}_\pm$  obtained by flowing from the critical orbits  $\mathcal{O}_\pm$ , can be derived from the curves that were depicted in fig. 4, but now in the  $x^2$  plane. On the other hand, the third Lefschetz thimble  $\mathcal{J}_0$  is associated to a similar curve connecting the point  $x^2 = 0$  to infinity. By an analysis similar to that of section 3.4.2, one can determine the jumping that occurs between  $\epsilon > 0$  and  $\epsilon < 0$ . The change in  $\mathcal{J}_\pm$  is precisely the same as it is for the Airy function; what is new is that  $\mathcal{J}_0$  picks up a linear combination of  $\mathcal{J}_\pm$ . We leave the details to the reader.

### 3.5 Analytic Continuation In $k$ : A Prototype

We have discussed at great length the analytic continuation of the integral of an exponential  $\exp(i\lambda f)$  from real to complex values of  $\lambda$ . For a more precise analog of Chern-Simons theory, we should consider analytic continuation of an *integer* parameter (such as the Chern-Simons level  $k$ ) to complex values.

For an example, we consider the integral representation of the Bessel function:

$$I(\kappa, \lambda) = \frac{1}{2\pi i} \oint \frac{dz}{z} z^\kappa \exp(\lambda(z - z^{-1})), \quad (3.66)$$

with an integer  $\kappa$  that in our analysis of section 5 will correspond roughly<sup>13</sup> to the Chern-Simons level  $k$ .

We take the integration cycle to be the circle  $|z| = 1$ . Alternatively, we can write

$$I(\kappa, \lambda) = \frac{1}{2\pi i} \oint \frac{dz}{z} \exp(\mathcal{I}), \quad (3.67)$$

with

$$\mathcal{I} = \lambda(z - z^{-1}) + \kappa \ln z. \quad (3.68)$$

For real  $\lambda$  and integral  $\kappa$ , this integral is an oscillatory integral of the general type that we have been considering, except that the integration cycle is compact, so that analytic continuation in  $\lambda$  poses no problem. However, we would like to analytically continue the integral to complex values of  $\kappa$ . This is a rough analog of analytically continuing Chern-Simons gauge theory with respect to the level  $k$ .

---

<sup>13</sup>In that analysis,  $\lambda$  corresponds roughly to the index  $n$  of the colored Jones polynomial  $J_n(q)$ , so the analogy would be closer if also  $\lambda$  were constrained to be an integer.



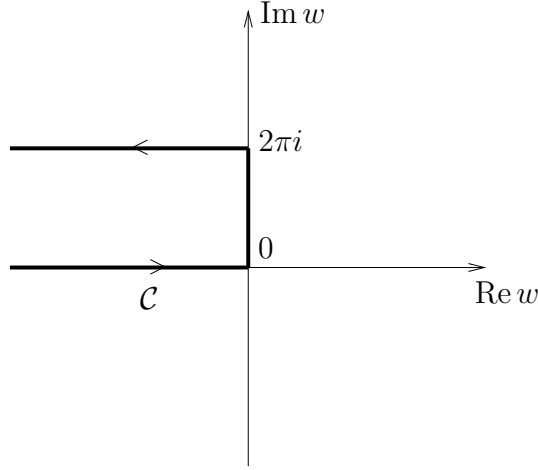


Figure 7: The contour  $\mathcal{C}$  in the complex  $t$ -plane that is used in analytic continuation of the Bessel function.

Let us recall the steps that were described somewhat abstractly in section 2.5.1. The holomorphic form that we are trying to integrate is naturally defined in the punctured  $z$ -plane, with the point  $z = 0$  omitted. To analytically continue in  $\kappa$ , we want to lift the integral from the punctured  $z$ -plane to its universal cover, which is the complex  $w$ -plane where  $z = e^w$ . The integration cycle  $|z| = 1$  can be lifted to the line segment in the  $w$ -plane from  $\alpha i$  to  $(2\pi + \alpha)i$ , for an arbitrary  $\alpha$ . Picking  $\alpha = 0$ , the integral becomes

$$I(\kappa, \lambda) = \frac{1}{2\pi i} \int_0^{2\pi i} dw \exp(2\lambda \sinh w + \kappa w). \quad (3.69)$$

The integration is now over an open contour whose ends are the two points  $w = 0$  and  $w = 2\pi i$ . If we stop here, Bessel's equation (whose proof requires integration by parts), will fail when  $\kappa \notin \mathbb{Z}$ . An analogous procedure in quantum field theory would lead to a failure of the Ward identities. Instead, following a classical procedure, and assuming that  $\text{Re } \lambda > 0$ , we add to the contour two more half lines given (fig.7) by the negative real axis and its translate by  $2\pi i$ , suitably oriented to give a closed although noncompact integration contour  $\mathcal{C}$ , which represents a cycle in the appropriate relative homology. The integral over this cycle

$$I(\kappa, \lambda) = \frac{1}{2\pi i} \int_{\mathcal{C}} dw \exp(2\lambda \sinh w + \kappa w) \quad (3.70)$$

converges (if  $\text{Re } \lambda > 0$ ), and in this integral,  $\kappa$  can now be treated as a complex variable. This gives an analytic continuation of the Bessel function to complex values of  $\kappa$ , since if  $\kappa$  is an integer, the two half lines that have been added to the integration contour cancel each other and the extended integral reduces to the original one. To analytically continue beyond the region  $\text{Re } \lambda > 0$ , one must shift  $\mathcal{C}$  vertically in the  $w$ -plane as the argument of  $\lambda$  varies. But we will mainly study the dependence on  $\kappa$  with  $\text{Re } \lambda > 0$ .

If  $\lambda$  and  $\kappa$  are real and obey suitable inequalities ( $\lambda$  and  $2\lambda + \kappa$  should be positive), the two half-lines that we added to complete the integration contour are precisely as described in section 2.5.1: they are the downward flow lines from the endpoints  $w = 0$  and  $w = 2\pi i$  of the naive integration contour in the complex  $w$  plane. If we vary  $\lambda$  and  $\kappa$  slightly, the flow lines move but without changing the relative homology class of the integration cycle  $\mathcal{C}$ .

We have described a simple procedure for analytic continuation away from integer values of  $\kappa$ , but there is no avoiding a basic fact discussed in section 2.5.2: analytic continuation away from integer values is not unique. For an illustrative special case of this, note that we could have begun by lifting the circle  $|z| = 1$  to a line segment in the  $w$ -plane connecting the points  $w = \pm\pi i$ . This would be analogous to a different choice of  $\alpha$  in section 2.5.1. Provided  $\lambda$  and  $\kappa$  are real and obey suitable inequalities ( $\lambda$  and  $2\lambda + \kappa$  should be negative), the downward flow lines from  $w = \pm i$  are again half-lines parallel to the real axis, but they extend in the opposite direction to  $\text{Re } w \gg 0$ . The two analytic continuations are qualitatively different. The original Bessel integral (3.66) obeys  $I(-\kappa, \lambda) = (-1)^\kappa I(\kappa, \lambda)$ , as follows from the change of variables  $z \leftrightarrow -1/z$ . This symmetry is broken by the choice of whether to extend the contour to  $\text{Re } w \ll 0$ , as in fig. 7, or to  $\text{Re } w \gg 0$ . With either choice, it turns out that when  $\kappa$  is not an integer, the large  $\kappa$  behavior depends very much on the sign of  $\text{Re } \kappa$ .

In the rest of our analysis, we will stick with the analytic continuation corresponding to the contour  $\mathcal{C}$  of fig. 7. We want to study the behavior of the integral (3.70) as  $\kappa$  and  $\lambda$  become large with a fixed value of  $\rho = \kappa/2\lambda$ . First we will learn what we can by inspection, taking  $\kappa$  real and  $\lambda$  positive, and then we will reconsider the matter using Morse theory.

The critical point equation is  $z^2 + 2\rho z + 1 = 0$  or  $\cosh w + \rho = 0$  and the critical points are at

$$z = -\rho \pm i\sqrt{1 - \rho^2}. \quad (3.71)$$

or

$$w = \pm \text{arccosh}(-\rho) + 2\pi i n = \ln(-\rho \pm \sqrt{\rho^2 - 1}) + 2\pi i n, \quad n \in \mathbb{Z}. \quad (3.72)$$

First let us suppose that  $\kappa$  is an integer, so that  $\mathcal{I}_{\kappa, \lambda}$  can be defined by the original integral on the contour  $|z| = 1$ , and we can use (3.71) to identify the critical points. If  $|\rho| \leq 1$ , the critical points are at  $|z| = 1$ , on the integration contour. They make oscillatory contributions to the integral, and this gives the behavior of the integral for large  $\kappa$  and fixed  $\rho$  in this region.

Now, still with integer  $\kappa$ , suppose that  $|\rho| > 1$ . The critical points move off the integration contour  $|z| = 1$ , and the integral therefore vanishes exponentially.

Next, suppose that  $\kappa$  is not an integer. We must consider the effects of the two half-lines that have been added to the integration contour. If  $|\rho| \leq 1$ , the addition of these

half-lines does not lead to a qualitative change; the integral is still dominated by oscillatory contributions from critical points at  $|z| = 1$ . We must be more careful if  $|\rho| > 1$ , and the two cases  $\rho > 1$  and  $\rho < -1$  turn out to be completely different.<sup>14</sup> If  $\rho > 1$ , there is no critical point on the integration contour and  $I(\kappa, \lambda)$  vanishes exponentially whether  $\kappa$  is an integer or not. (The most precise analysis is by a saddle point analysis that we describe shortly.) For  $\rho < -1$  (which means negative  $\kappa$ , since we have taken  $\lambda > 0$ ), the behavior is more delicate. There are two critical points on our chosen integration cycle  $\mathcal{C}$ , namely at  $w = \ln(-\rho - \sqrt{\rho^2 - 1})$  and  $w = \ln(-\rho - \sqrt{\rho^2 - 1}) + 2\pi i$ . The two critical points have the same positive value of  $h = \text{Re } \mathcal{I}$ , and they both make contributions that are exponentially large for  $\kappa \rightarrow -\infty$ . (Indeed,  $h \sim -\kappa f(\rho)$  for  $-\kappa$  large, where the function  $f(\rho) = (\rho^2 - 1)^{1/2}/\rho + \ln(-\rho + (\rho^2 - 1)^{1/2})$  is positive for  $\rho < -1$ .) On the other hand,  $\text{Im } \mathcal{I}$  vanishes for one critical point and is  $\kappa \text{Im } w = 2\pi\kappa$  at the other. Allowing for the opposite orientations of the two half-lines that are part of the contour  $\mathcal{C}$ , the two real critical points contribute

$$\exp(-\kappa f(\rho)) (1 - \exp(2\pi i \kappa)) \quad (3.73)$$

times the usual sort of asymptotic series. For integer  $\kappa$ , these contributions cancel, but for generic  $\kappa \ll 0$ , their sum grows exponentially with  $|\kappa|$ .

### 3.5.1 Description By Lefschetz Thimbles

Finally, we will seek a more precise description by expressing the cycle  $\mathcal{C}$  as a sum of Lefschetz thimbles  $\mathcal{J}_\sigma$  associated to critical points  $p_\sigma$ . Let us formulate the most parsimonious hypothesis that is consistent with the observations above:

(1) For  $\rho > 1$ , it is reasonable to guess that  $\mathcal{C}$  is equivalent to the Lefschetz thimble associated to the critical point at  $w = \ln(-\rho + \sqrt{\rho^2 - 1})$  ( $w$  has negative real part and we pick the branch of the logarithm such that  $\text{Im } w = \pi$ ). This is the unique critical point enclosed by  $\mathcal{C}$  if  $\rho > 1$ , and as  $h$  is negative at this critical point, the associated Lefschetz thimble makes an exponentially small contribution to the integral, as expected.

(2) For  $1 > \rho > -1$ , the two critical points at  $|z| = 1$  must certainly be included. There is no obvious need for more.

(3) For  $\rho < -1$ , the situation is more tricky. The two real critical points that account for the exponentially growing asymptotics in (3.73) must certainly contribute when  $\mathcal{C}$  is expressed in terms of Lefschetz thimbles. But as they cancel when  $\kappa$  is an integer, while  $I(\kappa, \lambda)$  is not zero in that case, the expression for  $\mathcal{C}$  in terms of Lefschetz thimbles must

---

<sup>14</sup>This difference results entirely from the choice of how to analytically continue the function  $I(\kappa, \lambda)$  to nonintegral values of  $\kappa$ , since for  $\kappa \in \mathbb{Z}$ , the Bessel function has a symmetry  $I(-\kappa, \lambda) = (-1)^\kappa I(\kappa, \lambda)$ , as already noted.

receive a contribution from at least one more critical point. This critical point must have  $h < 0$ , since  $I(\kappa, \lambda)$  is exponentially small for integer  $\kappa$  and  $\rho < -1$ .

In summary, in regions (1), (2), or (3),  $\mathcal{C}$  must be a sum of at least 1, 2, or 3 critical point contours, respectively. This most parsimonious interpretation is actually correct, as one can see from fig.8, where the relevant Lefschetz thimbles are sketched.

### 3.5.2 Singularities At $\rho = \pm 1$

The most interesting values of  $\rho$  are at  $\rho = \pm 1$ , where critical points meet. Let us first examine the behavior at these points on the  $z$ -plane, and then on the  $w$ -plane. The analysis on the  $z$ -plane is simpler, but it only tells us what happens when  $\kappa$  is an integer. To analyze the behavior for non-integral  $\kappa$ , we need to work on the  $w$ -plane.

There are precisely two critical points on the  $z$ -plane, located at  $z = -\rho \pm i\sqrt{1 - \rho^2}$ . Precisely for  $\rho = \pm 1$ , the two critical points meet. This is reminiscent of the most basic example of critical behavior, the cubic polynomial  $f(x) = x^3/3 - \epsilon x$  related to the Airy function, whose two critical points meet at  $\epsilon = 0$ . The critical points of  $f$  are real for  $\epsilon > 0$  and form a complex conjugate pair for  $\epsilon < 0$ . The consequences of the singularity at  $\epsilon = 0$  were explored in section 3.4.1.

Our problem on the  $z$ -plane is exactly analogous. We endow the punctured  $z$ -plane with the antiholomorphic involution  $z \rightarrow 1/\bar{z}$ , so that the real points (the fixed points of the involution) are the circle  $|z| = 1$ . The two critical points at  $z = -\rho \pm i\sqrt{1 - \rho^2}$  are real for  $\rho$  real and  $|\rho| < 1$  but form a complex conjugate pair for  $\rho$  real and  $|\rho| > 1$ . The singularities at  $\rho = \pm 1$  are just analogous to the Airy function singularity at  $\epsilon = 0$ .

As long as  $\kappa$  is an integer and we work on the  $z$ -plane, the behavior of the Bessel integral that we found in section 3.5.1 precisely mirrors the behavior of the Airy integral

$$I(\epsilon) = \int_{-\infty}^{\infty} dx \exp(i(x^3/3 - \epsilon x)). \quad (3.74)$$

Region (2) of the Bessel integral corresponds to the case  $\epsilon > 0$  for the Airy integral: the integration cycle is equivalent to the sum of two Lefschetz thimbles associated to two real critical points. Regions (1) and (3) of the Bessel integral correspond to the case  $\epsilon < 0$  for the Airy integral: the integral is equivalent to an integral over a single Lefschetz thimble that makes an exponentially small contribution. In section 3.5.1, we found that three critical points are needed in region (3) to describe the integral on the  $w$ -plane, but two of them (the ones that make exponentially growing contributions) cancel upon projection to the  $z$ -plane.

The analysis on the  $w$ -plane is more complicated because each critical point on the  $z$ -

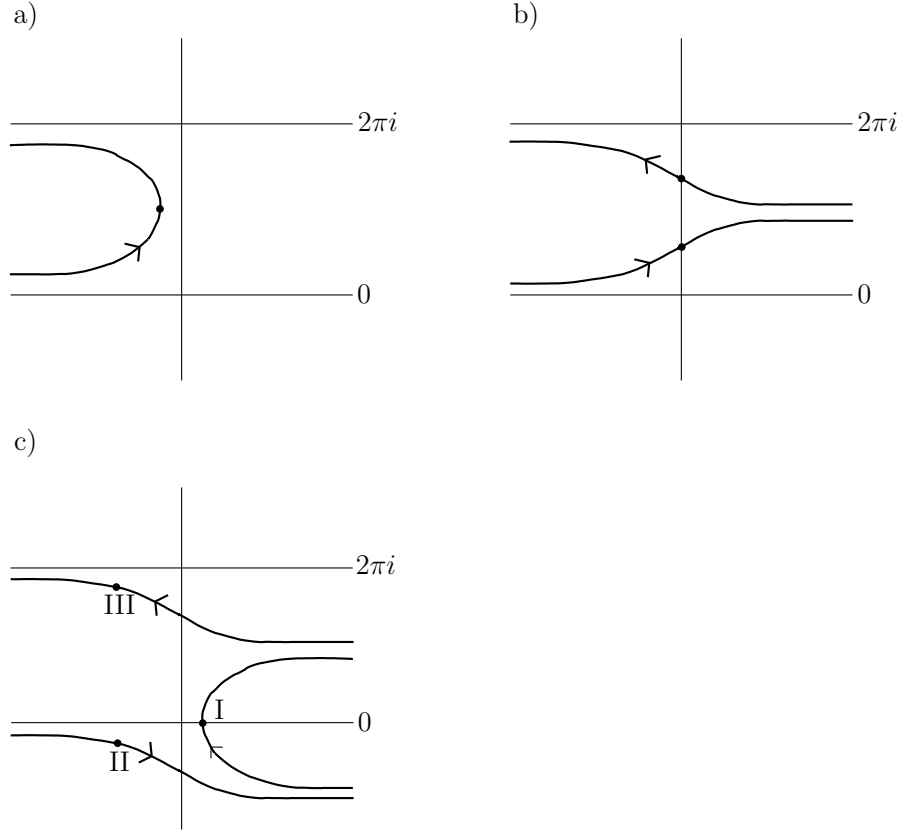


Figure 8: Qualitative behavior of Lefschetz thimbles in analytic continuation of the Bessel function, for (a)  $\rho > 1$ , (b)  $1 > \rho > -1$ , and (c)  $-1 > \rho$ . In each case, the critical points are marked by black dots. In (a), the integration cycle of fig. 7 is equivalent to a single Lefschetz thimble associated to a complex critical point that makes an exponentially decaying contribution to the integral. In (b), it is equivalent to the sum of two Lefschetz thimbles that make oscillatory contributions. And in (c), it is equivalent to the sum of a Lefschetz thimble associated to a critical point (labeled as I) that makes an exponentially decaying contribution and two critical points (labeled II and III) whose contributions are exponentially growing, but cancel if  $\kappa$  is an integer. In (c),  $\rho$  has been given a small imaginary part to avoid a Stokes line.

plane lifts to infinitely many critical points on the  $w$ -plane (differing by  $w \rightarrow w + 2\pi in$ ,  $n \in \mathbb{Z}$ ). Nevertheless, it is very illuminating to repeat the analysis of the Bessel function emphasizing its interpretation in terms of Airy function singularities:

(1') We start for  $\rho > 1$  with a single critical point at  $\tilde{w}_- = \ln(-\rho + \sqrt{\rho^2 - 1})$ , picking the branch of the logarithm so that  $\tilde{w}_- = \pi i$  at  $\rho = 1$ . (Choosing a different branch would multiply the analytically continued Bessel integral by an integer power of  $\exp(2\pi i \kappa)$ .) The name  $\tilde{w}_-$  is motivated by the fact that in section 3.4.1, a complex critical point that makes an exponentially small contribution to the Airy function for  $\epsilon < 0$  was called  $\tilde{p}_-$ . As in section 3.4.1, we write  $\tilde{\mathcal{J}}_-$  for the Lefschetz thimble corresponding to this critical point. For  $\rho > 1$ , as we have learned in section 3.5.1, the integration cycle  $\mathcal{C}$  that we have used to analytically continue the Bessel function is equivalent to  $\tilde{\mathcal{J}}_-$ .

(2') The  $\rho > 1$  critical point at  $\tilde{w}_- = \ln(-\rho + \sqrt{\rho^2 - 1})$ , when continued to  $\rho = 1$ , meets two real critical points that for  $\rho < 1$  are located at  $w_{\pm} = \ln(-\rho \pm i\sqrt{1 - \rho^2})$ . Here branches of the logarithm are chosen so that at  $\rho = 1$ , we have  $w_+ = w_- = \tilde{w}_- = \pi i$ . As in the Airy case, we denote as  $\mathcal{J}_+$  and  $\mathcal{J}_-$  the Lefschetz thimbles that correspond to the real critical points  $w_{\pm}$ . We can use the Airy relation (3.58) to express the region (1) integration cycle  $\tilde{\mathcal{J}}_-$  in terms of Lefschetz thimbles  $\mathcal{J}_{\pm}$  appropriate to region (2). The relation is  $\tilde{\mathcal{J}}_- = \mathcal{J}_+ + \mathcal{J}_-$ . So the right integration cycle in region (2) is  $\mathcal{J}_+ + \mathcal{J}_-$ .

(3') So far working on the  $w$ -plane instead of the  $z$ -plane has not given anything essentially new. Now we study how the critical points  $w_{\pm}$  evolve as we vary  $\rho$  from 1 to  $-1$ . We have chosen branches of the logarithm so that  $w_{\pm} = \ln(-\rho \pm i\sqrt{1 - \rho^2})$  are both equal to  $\pi i$  at  $\rho = 1$ . However, when continued (along the real  $\rho$  axis) to  $\rho = -1$ , they are not equal. Rather,

$$w_+(\rho = -1) = 0, \quad w_-(\rho = -1) = 2\pi i. \quad (3.75)$$

Hence, we cannot analyze what is happening at  $\rho = -1$  in terms of a single Airy singularity. We must consider two different Airy singularities, located at  $w = 0$  and at  $w = 2\pi i$ . This will make the analysis a little longer, though the basic idea is just to use what we know about the Airy integral.

Near  $w = 0$ , let us write  $\mathcal{J}'_{\pm}$  for the Lefschetz thimbles associated to real critical points at  $\rho > -1$ , and  $\tilde{\mathcal{J}}'_{\pm}$  for the Lefschetz thimbles associated to complex critical points at  $\rho < -1$ . We write  $\mathcal{J}''_{\pm}$  and  $\tilde{\mathcal{J}}''_{\pm}$  for Lefschetz thimbles associated to the analogous real and complex critical points near  $w = 2\pi i$ . (The reason for the notation  $\mathcal{J}'$  and  $\mathcal{J}''$  is to distinguish these cycles from the ones we have already defined near  $\rho = 1$ .)

We have already found that for  $\rho$  just less than 1, the integration cycle is  $\mathcal{J}_+ + \mathcal{J}_-$ . Continuing along the real axis to  $\rho$  just greater than  $-1$  with the aid of eqn. (3.75), we

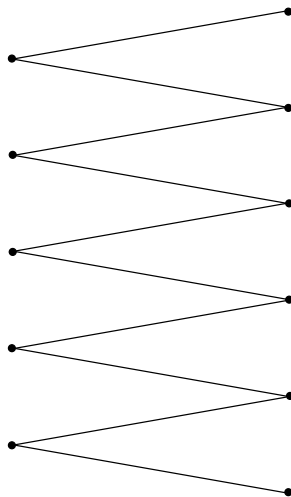


Figure 9: Qualitative behavior of the conserved quantity  $\text{Im } I/\kappa$  for critical points in the region  $-1 \leq \rho \leq 1$ .  $\rho$  is plotted horizontally and  $\text{Im } \mathcal{I}/\kappa$  vertically. A black dot represents a pair of real critical points that become coincident at  $\rho = 1$  or  $\rho = -1$ . At  $\rho = 1$  (right), there are a pair of real critical points for every value of  $\text{Im } \mathcal{I}/\kappa$  of the form  $(2w + 1)\pi$ ,  $w \in \mathbb{Z}$ . Reducing  $\rho$ ,  $\text{Im } \mathcal{I}/\kappa$  increases for one of these critical points and reduces for the other. At  $\rho = -1$ , the critical points recombine in pairs, but now with  $\text{Im } \mathcal{I}/\kappa$  of the form  $2\pi w$ ,  $w \in \mathbb{Z}$ . In all cases, these critical points are found by solving the equation  $\cosh w = -\rho$ .

see that  $\mathcal{J}_+ + \mathcal{J}_-$  evolves to  $\mathcal{J}'_+ + \mathcal{J}''_-$ . To continue past  $\rho = -1$ , we use the Airy relations (3.57), which give  $\mathcal{J}'_+ = -\tilde{\mathcal{J}}'_+ + \tilde{\mathcal{J}}'_-$ , and  $\mathcal{J}''_- = \tilde{\mathcal{J}}''_+$ . So  $\mathcal{J}'_+ + \mathcal{J}''_- = -\tilde{\mathcal{J}}'_+ + \tilde{\mathcal{J}}''_+ + \tilde{\mathcal{J}}'_-$ , and this is the integration cycle for  $\rho < -1$ . For integer  $\kappa$ ,  $-\tilde{\mathcal{J}}'_+$  and  $\tilde{\mathcal{J}}''_+$  cancel and the integration cycle is equivalent to  $\tilde{\mathcal{J}}'_-$ . For general  $\kappa$ , the contribution from  $-\tilde{\mathcal{J}}'_+ + \tilde{\mathcal{J}}''_+$  dominates and is described qualitatively in eqn. (3.73).

With a view to our application in section 5.3, the key point to remember is that the two real critical points on the  $z$ -plane, if lifted to the  $w$ -plane so that they coincide at  $\rho = 1$ , do not coincide at  $\rho = -1$ . The flow of critical points is sketched in fig. 9.

## 4 Application To Chern-Simons Theory

Can the framework of section 3 be applied in the infinite-dimensional context of Chern-Simons gauge theory? The critical points  $p_\sigma$  certainly make sense. The Euler-Lagrange equations simply say that the gauge field should be flat (perhaps when restricted to the complement of a knot). Moreover, by standard quantum field theory methods, one certainly knows how to expand the path integral around any given critical point, and hence in effect

the Lefschetz thimbles  $\mathcal{J}_\sigma$  are known at least perturbatively.

The aspects of this framework that are not standard in quantum field theory are the use of the flow equation to define the critical point cycles nonperturbatively, to determine the coefficients when the real integration cycle is expressed in terms of critical point cycles, making analytic continuation possible, and to describe the Stokes jumping phenomena that are essential for the consistency of the framework, once one aims for nonperturbative results. All of these questions involve the flow equations, so the question of whether the Morse theory machinery can be applied to Chern-Simons theory is largely the question of whether the flow equation for Chern-Simons makes sense and has properties similar to what it has in finite dimensions.

This is the topic of section 4.1. The flow equations in ordinary Morse theory are ordinary differential equations that describe flow in an auxiliary “time” direction. Chern-Simons theory is three-dimensional even before introducing an auxiliary direction, so the flow equations become partial differential equations in four dimensions. Prototypes for such equations are the Yang-Mills instanton equations, which are interpreted in Floer theory as flow equations for a real connection  $A$  with the Chern-Simons function  $W(A)$  as a Morse function. In that example, the flow equation has a four-dimensional symmetry, and this is a hint of what we will find in section 4.1.

Our problem is analogous to the case just mentioned but is more elaborate because we want the flow for a complex connection  $\mathcal{A}$  with a Morse function derived from the complex Chern-Simons function  $W(\mathcal{A})$ . Also our flow equations depend on a coupling parameter. It turns out that the flow equations in our problem are essentially equations that have appeared before, in the application of supersymmetric gauge theory to the geometric Langlands program [53]. The reason for this is not yet fully clear, but at any rate the elliptic nature of the flow equations for Chern-Simons theory means that they have properties similar to the Morse theory flow equations with finitely many variables.

## 4.1 The Flow Equation For Chern-Simons

Let  $M$  be a three-manifold. Let  $H$  be a compact Lie group and set  $G = H_{\mathbb{C}}$ . Let  $\mathcal{T}$  be the space of connections  $A$  on a fixed  $H$ -bundle  $E \rightarrow M$ , and let  $\mathcal{T}_{\mathbb{C}}$  be the space of complex-valued connections  $\mathcal{A}$  on the complexification of  $E$ . We can write  $\mathcal{A} = A + i\phi$  where  $A$  and  $\phi$  are the real and imaginary parts of  $\mathcal{A}$ ;  $A$  is an  $\mathfrak{h}$ -valued connection, and  $\phi$  is a one-form on  $M$  with values in the adjoint bundle  $\text{ad}(E)$ . The curvature of  $A$  will be denoted as  $F = dA + A \wedge A$ , and that of  $\mathcal{A}$  as  $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ .

$\mathcal{T}_{\mathbb{C}}$  is a complex manifold; to define a flow equation, we need to endow it with a Kahler



metric. We pick a Riemannian metric  $g$  on  $M$ , which determines a Hodge star operator<sup>15</sup>  $\star_M$ . We define the metric on  $\mathcal{T}$  by

$$ds^2 = - \int_M \text{Tr } \delta \overline{\mathcal{A}} \wedge \star_M \delta \mathcal{A}. \quad (4.1)$$

This is a Kahler metric, with Kahler form

$$\omega = \int_M \text{Tr } \delta \phi \wedge \star_M \delta A. \quad (4.2)$$

The metric  $ds^2$  is invariant under  $H$ -valued local gauge transformations, though not under  $G$ -valued ones. The moment map for the  $H$ -valued local gauge transformations is

$$\mu = d_A \star_M \phi, \quad (4.3)$$

where  $d_A = d + [A, \cdot]$  is the covariant derivative with respect to  $A$ .

Critical “points” of the Chern-Simons function  $W(\mathcal{A})$ , understood as a function on the space of connections, actually form orbits under the group of complex gauge transformations. For reasons described in section 3.3.3, we will only consider critical orbits that are semistable, meaning that they admit a locus with  $\mu = 0$ . The condition for a flat connection  $\mathcal{A}$ , that is a connection obeying  $0 = \mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ , to be gauge-equivalent to a connection for which  $\mu = 0$  is [52] that the holonomy of the flat connection  $\mathcal{A}$  is not strictly triangular. In other words, if the holonomies of  $\mathcal{A}$  around one-cycles in  $M$  can be simultaneously put in upper triangular form

$$\begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix}, \quad (4.4)$$

then the upper right elements  $\beta$  all vanish. Strictly triangular flat connections are unstable and we omit them. (A block diagonal flat connection with  $\beta = 0$  is considered semistable, and one that cannot be put in a triangular form at all is stable.)

The flow equation for a Morse function that is the real part of a holomorphic function  $\mathcal{I}$  is

$$\frac{d\mathcal{A}}{dt} = - \star_M \frac{\delta \mathcal{I}}{\delta \overline{\mathcal{A}}}. \quad (4.5)$$

We are interested in the case that  $\mathcal{I} = \lambda W(\mathcal{A})$  where  $W$  is the Chern-Simons functional and  $\lambda$  is a complex number. By rescaling  $t$ , we can reduce to the case that  $\lambda = 2\pi \exp(-i\alpha)$  with real  $\alpha$ . The flow equation then becomes

$$\frac{d\mathcal{A}}{dt} = - \exp(-i\alpha) \star_M \overline{\mathcal{F}}, \quad (4.6)$$

---

<sup>15</sup> If  $dx^1, dx^2, dx^3$  is an oriented orthonormal frame, we pick the sign of  $\star_M$  so that  $\star_M(dx^1) = dx^2 \wedge dx^3$ , etc. Similarly, later the star operator of  $I \times M$  is defined so that  $\star(dt \wedge dx^1) = dx^2 \wedge dx^3$ , etc.

where  $\overline{\mathcal{F}}$  is the complex conjugate of the curvature  $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ . More explicitly, this is

$$\begin{aligned}\frac{dA}{dt} &= -\star_M (\cos \alpha (F - \phi \wedge \phi) - \sin \alpha d_A \phi) \\ \frac{d\phi}{dt} &= \star_M (\sin \alpha (F - \phi \wedge \phi) + \cos \alpha d_A \phi).\end{aligned}\tag{4.7}$$

The time  $t$  parametrizes a real one-manifold  $I$  (which is the real line or a portion of it). We pull back  $E$  to a bundle over  $I \times M$  and view  $A$  as a connection on this pullback. As such,  $A$  has vanishing component in the  $I$  direction. We view this as a gauge condition. Relaxing this gauge condition, we can write  $dA_i/dt$  ( $i$  is an index tangent to  $M$ ) in gauge-invariant language as a component  $F_{ti}$  of the curvature  $F = dA + A \wedge A$ . Similarly, we view  $\phi$  as an  $\text{ad}(E)$ -valued one-form on  $I \times M$ ; as such, it is not generic, since its component  $\phi_t$  in the  $I$  direction vanishes. We endow  $I \times M$  with a Riemannian metric that is the sum of the metric  $dt^2$  on  $I$  with the metric  $g$  on  $M$ , and we write simply  $\star$  for the Hodge star operator on  $I \times M$  (normalized as in footnote 15). After taking linear combinations of the equations in (4.7), one finds that these equations can be put in a form with four-dimensional symmetry

$$\begin{aligned}(F - \phi \wedge \phi)^+ &= u D\phi^+ \\ (F - \phi \wedge \phi)^- &= -u^{-1} D\phi^-, \end{aligned}\tag{4.8}$$

with

$$u = \frac{1 - \cos \alpha}{\sin \alpha}, \quad u^{-1} = \frac{1 + \cos \alpha}{\sin \alpha}.\tag{4.9}$$

(For a two-form  $v$ , we define the projections  $v^\pm = (v \pm \star v)/2$ .) To be more precise, the four-dimensional symmetry is broken in our context by the condition  $\phi_t = 0$ . But if we relax this condition, the equations have full four-dimensional symmetry.

The moment map  $\mu = d_A \star_M \phi$  is as usual a conserved quantity for the flow equations. As explained in section 3.3.3, we are primarily interested in solutions of the flow equations with  $\mu = 0$ . Bearing in mind that we anyway have the condition  $\phi_t = 0$ ,  $d_A \star_M \phi$  can be written in four-dimensional language as  $d_A \star \phi$ . So we want to supplement the equations (4.8) with an additional condition

$$d_A \star \phi = 0.\tag{4.10}$$

Henceforth, the phrase “flow equations” will refer to this combined system of equations.

The flow equations have been first encountered [53] in the gauge theory approach to geometric Langlands.<sup>16</sup> These are elliptic equations, modulo the gauge symmetry, and this means that their general properties are much like those of flow equations for finitely many variables in ordinary Morse theory. The relevant properties are stated in section (4.1.2).

---

<sup>16</sup>The parameter  $u$  is called  $t$  in [53], but here we prefer to avoid confusion with the time coordinate.

The condition  $\phi_t = 0$  can be reinterpreted in terms of a vanishing theorem similar to those of section 3.3 of [53]. This argument proceeds as follows. Set  $\mathcal{V}^+(u) = (F - \phi \wedge \phi)^+ - uD\phi^+$ ,  $\mathcal{V}^-(u) = (F - \phi \wedge \phi)^- + u^{-1}D\phi^-$ ,  $\mathcal{V}_0 = d_A \star \phi$ , so the flow equations are  $\mathcal{V}^+(u) = \mathcal{V}^-(u) = \mathcal{V}_0 = 0$ . Now consider the following sum of squares of these equations:

$$\begin{aligned} & - \int_{I \times M} \text{Tr} \left( \frac{u^{-1}}{u + u^{-1}} \mathcal{V}^+(u) \wedge \mathcal{V}^+(u) + \frac{u}{u + u^{-1}} \mathcal{V}^-(u) \wedge \mathcal{V}^-(u) + \mathcal{V}_0 \wedge \star \mathcal{V}_0 \right) \\ & = - \int_{I \times M} d^4x \sqrt{g} \text{Tr} \left( \frac{1}{2} F_{\mu\nu} F^{\mu\nu} + D_\mu \phi_\nu D^\mu \phi^\nu + R_{\mu\nu} \phi^\mu \phi^\nu + \frac{1}{2} [\phi_\mu, \phi_\nu]^2 \right) \\ & \quad + \frac{u - u^{-1}}{u + u^{-1}} \int_{I \times M} \text{Tr} F \wedge F. \end{aligned} \tag{4.11}$$

(Indices  $\mu, \nu$  are tangent to  $I \times M$ , and  $R_{\mu\nu}$  is the Ricci tensor of  $I \times M$ , which is a pullback from  $M$ .) Clearly, the left hand side of (4.11) is stationary when the flow equations are obeyed. Therefore the right hand side is also stationary. The part of the right hand side that depends on  $\phi_t$  is simply

$$Z(\phi_t) = - \int_{I \times M} d^4x \sqrt{g} \text{Tr} \left( \sum_\mu D_\mu \phi_t D^\mu \phi_t + \sum_{\mu \neq t} [\phi_t, \phi_\mu]^2 \right). \tag{4.12}$$

The flow equations therefore imply that  $\phi_t$  obeys the Euler-Lagrange equations derived from (4.12), or

$$- \sum_\mu D_\mu D^\mu \phi_t + \sum_{\mu \neq t} [\phi_\mu, [\phi_\mu, \phi_t]] = 0. \tag{4.13}$$

(Of course, this can also be proved by differentiating the equations (4.8) and (4.10).) Provided that one can integrate by parts, which is true if  $\phi_t$  obeys either Dirichlet or Neumann boundary conditions at each end of  $I$ , (4.13) implies that  $Z(\phi_t) = 0$ . Given the definition of  $Z(\phi_t)$  as a sum of squares, this immediately implies that  $\phi_t$  is covariantly constant and commutes with all  $\phi_\mu$ . Thus,  $\phi_t$  generates a symmetry of the solution.

We will always choose boundary conditions such that  $\phi_t$  vanishes at (at least) one end of  $I$ , and integration by parts is possible at the other end. The above argument shows that in a solution of the flow equations obeying such boundary conditions,  $\phi_t$  is identically zero. There is no need to spoil the four-dimensional symmetry, or the elliptic nature of the equations, by postulating in advance that  $\phi_t$  is identically zero. This is a consequence of the equations plus boundary conditions.

#### 4.1.1 The Index

For applications, we need to know the expected value of the dimension of the moduli space  $\mathcal{M}$  of solutions of the above equations, with appropriate boundary conditions and modulo

gauge transformations. This equals the index of the linear elliptic operator  $L$  that arises from those equations, after linearization and gauge fixing. We will show that on a four-manifold of the form  $I \times M$ , with certain types of boundary condition at the ends of  $I$ , the index vanishes. This will lead in section 4.1.2 to behavior similar to that of Morse theory with finitely many variables.

The index is independent of  $u$ , so it suffices to set  $u = 1$ . The relevant linear system can be described as follows. We expand around a classical solution  $(A_0, \phi_0)$  by setting  $(A, \phi) = (A_0 + \delta A, \phi_0 + \delta \phi)$ . A convenient gauge-fixing condition is  $d_{A_0} \star \delta A = 0$ . For a four-manifold  $X$ , let  $\Omega_X^q(\text{ad}(E))$  be the space of  $q$ -forms on  $X$  with values in  $\text{ad}(E)$ .

After acting on  $\delta \phi$  with the  $\star$  operator, we can think of  $\delta A$  and  $\delta \phi$  as elements of  $\Omega_X^1(\text{ad}(E))$  and  $\Omega_X^3(\text{ad}(E))$ , respectively. We can view the gauge-fixing condition as asserting the vanishing of the element  $\star d_{A_0} \star \delta A$  of  $\Omega_X^0(\text{ad}(E))$ , while the equations  $\mathcal{V}^+ = \mathcal{V}^- = 0$  assert the vanishing of the element  $\mathcal{V}^+ + \mathcal{V}^-$  of  $\Omega_X^2(\text{ad}(E))$ , and the equation  $\mathcal{V}_0 = 0$  asserts the vanishing of an element of  $\Omega_X^4(\text{ad}(E))$ .

All told, our linear operator  $L$  maps  $\Omega_X^1(\text{ad}(E)) \oplus \Omega_X^3(\text{ad}(E))$  to  $\Omega_X^0(\text{ad}(E)) \oplus \Omega_X^2(\text{ad}(E)) \oplus \Omega_X^4(\text{ad}(E))$ . A familiar operator that does this is the operator  $L_0 = d_A + d_A^*$  mapping  $\text{ad}(E)$ -valued differential forms of odd degree to those of even degree.  $L$  is actually equal to  $L_0$  modulo zeroth order terms that do not affect the index. As we explain momentarily, the index of the operator  $L_0$ , or of  $L$ , is unaffected by the twisting by the gauge bundle  $E$ . If  $X$  is compact, this index is simply the product  $-\chi(X) \dim(H)$ , where  $\chi(X)$  is the Euler characteristic of  $X$  and  $\dim(H)$  is the dimension of the Lie group  $H$ . For  $X = I \times M$ , the topological Euler characteristic of  $X$  vanishes, and boundary contributions to the index are also zero, so the index of  $L$  vanishes.

The value of the index for compact  $X$  and its vanishing for  $X = I \times M$  can be understood as follows. Let  $S_+$  and  $S_-$  be the two spin bundles of  $X$ ; each is of rank 2. It is convenient to express the operator  $L_0$  as a Dirac operator mapping sections of  $S_+ \otimes V$  to sections of  $S_- \otimes V$  for some vector bundle or difference of vector bundles  $V$ ; the index of  $L_0$  can then be computed from the index theorem for the Dirac operator. To put  $L_0 = d_A + d_A^*$  in the desired form, we must take  $V$  to be a formal difference of vector bundles  $V = (S_+ \ominus S_-) \otimes \text{ad}(E)$ . Taking the formal difference  $S_+ \ominus S_-$  where  $S_+$  and  $S_-$  have the same rank causes all bulk and boundary contributions to the index that depend on  $E$  to cancel. (As  $X$  is a four-manifold and  $E$  has vanishing first Chern class, the only bulk  $E$ -dependent contribution is a multiple of the second Chern class of  $E$ ; it cancels when we take the difference  $S_+ \ominus S_-$ . Boundary contributions are further discussed in the next paragraph.) Hence  $\text{ad}(E)$  can be replaced by a trivial bundle of the same rank, namely  $\dim(H)$ , and  $L_0$  then reduces to the sum of that number of copies of the standard operator  $d + d^*$  mapping ordinary differential forms of odd degree to those of even degree. The index of this operator on a four-manifold without boundary is  $-\chi(X)$ .

Since the vanishing of boundary corrections to the index is important, let us spell it out in more detail. If  $I$  has a boundary at a finite distance – for example the right boundary of the half-line  $(-\infty, 0]$  – then we will pick local boundary conditions. With local boundary conditions, there is no  $E$ -dependent boundary contribution to the index, and as the metric of  $I \times M$  is a product, there is also no boundary contribution involving the extrinsic curvature of the boundary. Alternatively, if  $I$  has an “end” at infinity – for example, the left end of the half-line – then we want normalizable solutions of the equation  $L\Psi = 0$ . Asking for normalizable solutions is equivalent to cutting off the half-line at a finite distance and imposing global Atiyah-Patodi-Singer boundary conditions. In general, in such a situation there is a boundary contribution to the index of a Dirac-like operator, but it cancels when we take the formal difference  $(S_+ \ominus S_-) \otimes \text{ad}(E)$  because  $S_+$  and  $S_-$  are equivalent when restricted to  $M$ .

The operator  $L$  is real, so it has a real determinant line bundle. This determinant line bundle is topologically trivial<sup>17</sup> and after fixing a trivialization,  $L$  has a natural real-valued determinant function  $\det L$ .

#### 4.1.2 Applications Of The Flow Equation

The flow equation has two primary applications. The first is to determine the coefficients when a given cycle  $\mathcal{C}$  in the relative homology is expanded as a sum of Lefschetz thimbles  $\mathcal{J}_\sigma$  in the usual way,

$$\mathcal{C} = \sum_{\sigma} \mathbf{n}_{\sigma} \mathcal{J}_{\sigma}. \quad (4.14)$$

We are only interested in cycles  $\mathcal{C}$  that lie at  $\mu = 0$ .

The coefficient  $\mathbf{n}_{\sigma}$  is obtained by counting, with suitable signs, the solutions of the upward flow equation on the half-line  $\mathbb{R}^- = (-\infty, 0]$ . As usual, the upward flow equations are obtained from the downward ones by reversing the sign of the time. The boundary conditions are that the flow starts at a flat connection  $\mathcal{A}_{\sigma}$  (corresponding to the cycle  $\mathcal{J}_{\sigma}$ ) at  $t = -\infty$  and ends on  $\mathcal{C}$  at  $t = 0$ . Equivalently, we can consider downward flow equations on the half-line  $\mathbb{R}^+ = [0, \infty)$ . In this case, we consider flows that start on  $\mathcal{C}$  at  $t = 0$  and end at  $\mathcal{A}_{\sigma}$  at  $t = \infty$ .

We write  $\mathcal{A}_M = A_M + i\phi_M$  for the restriction of  $\mathcal{A} = A + i\phi$  to  $M$ , at a fixed  $t$ . The boundary condition that we want at  $t = 0$  is that  $\mathcal{A}_M$  should lie in  $\mathcal{C}$ . Since we assume that  $\mathcal{C}$

---

<sup>17</sup>In Donaldson theory, it is shown that the determinant line bundle of the Dirac operator twisted by  $S_+ \otimes \text{ad}(E)$  is trivial under a mild topological restriction that is satisfied in our applications. In our case, because we are twisting by the difference  $(S_+ \ominus S_-) \otimes \text{ad}(E)$ , the topological condition in question is not needed.

lies at  $\mu = 0$ , the condition that  $\mathcal{A}_M$  lies in  $\mathcal{C}$  implies in particular that  $\mu = d_A \star_M \phi_M = 0$  at  $t = 0$ . Subtracting this from the bulk equation  $d_A \star \phi = 0$ , which is one of the flow equations, we find that the covariant normal derivative of  $\phi_t$  vanishes at  $t = 0$ :  $D_t \phi_t = 0$ . This is enough to justify integration by parts in showing that  $Z(\phi_t)$ , defined in (4.12), vanishes. So it will lead to global vanishing of  $\phi_t$  after we also pick a suitable boundary condition at  $t = -\infty$ .

The most basic example is the case that  $\mathcal{C}$  is the real integration cycle of Chern-Simons theory with compact gauge group  $H$ . In this case, the condition for  $\mathcal{A}_M$  to lie in  $\mathcal{C}$  is simply  $\phi_M = 0$ . This certainly implies that  $d_A \star \phi_M = 0$ .

At  $t = -\infty$  (or  $t = +\infty$  if we formulate the flow on  $\mathbb{R}^+$ ), we ask that  $\mathcal{A}_M$  approaches a specified flat connection  $\mathcal{A}_\sigma$  and that  $\phi_t$  vanishes. Given this vanishing, and the fact that we have chosen a boundary condition that allows integration by parts at  $t = 0$ , the vanishing theorem implies that  $\phi_t$  will vanish everywhere.

Let  $\mathcal{M}$  be the moduli space of solutions of the flow equations with boundary conditions just described. Since the index in the linearized problem vanishes, the expected dimension of  $\mathcal{M}$  is zero and for a generic metric on  $M$ ,  $\mathcal{M}$  consists of a finite set of points. We count these points with a weight that is 1 or  $-1$  depending on the sign<sup>18</sup> of  $\det L$ . Summing the contributions of all critical points, we get the integer  $\mathbf{n}_\sigma$  that appears in the expansion  $\mathcal{C} = \sum_\sigma \mathbf{n}_\sigma \mathcal{J}_\sigma$ .

Let us compare the definition of  $\mathbf{n}_\sigma$  to the definition of the most basic invariant in Donaldson theory of smooth four-manifolds. This invariant is defined in a situation in which the index of the operator  $L_-$  that arises in linearizing the Yang-Mills instanton equation on  $X$  is zero. Under this circumstance, for a generic metric on  $X$ , the instanton equation has finitely many solutions, and Donaldson's invariant is defined by summing over all such solutions with a weight given by the sign of  $\det L_-$ . Obviously,  $\mathbf{n}_\sigma$  is a close cousin of this, with the usual instanton equations replaced by the flow equations for a complex-valued connection.

So far, the elliptic nature of the flow equations has enabled us to express a given cycle  $\mathcal{C}$  in terms of the Lefschetz thimbles  $\mathcal{J}_\sigma$ , according to eqn. (4.14). The other important application of the flow equation is to describe Stokes phenomena. Here we must consider complete flow lines between two critical points corresponding to semistable flat connections  $\mathcal{A}_\sigma$  and  $\mathcal{A}_\tau$  on  $M$ . We must define an invariant  $\mathbf{m}_{\sigma\tau}$  that controls what happens in crossing a Stokes ray at which there are flows from  $\mathcal{A}_\sigma$  to  $\mathcal{A}_\tau$ .

We consider the flow equations on  $X = \mathbb{R} \times M$ , with the requirement that  $\mathcal{A}_M \rightarrow \mathcal{A}_\sigma$  for

---

<sup>18</sup>The overall sign of the function  $\det L$  depends on a choice of orientation of the relevant determinant line bundle. This amounts to a choice of relative orientations of the cycles  $\mathcal{C}$  and  $\mathcal{J}_\sigma$  that we are comparing. Such choices of sign and orientation are discussed in, for example, [46] in the context of finite-dimensional Morse theory and Floer cohomology.

$t \rightarrow -\infty$  and  $\mathcal{A}_M \rightarrow \mathcal{A}_\tau$  for  $t \rightarrow +\infty$ . Further we require that  $\phi_t \rightarrow 0$  at both ends. The index of  $L$  still vanishes, but now there is a new ingredient: the flow equations are invariant under the group  $\mathbf{T} \cong \mathbb{R}$  of translations in the  $\mathbb{R}$  direction, but a solution with  $\sigma \neq \tau$  cannot be translation-invariant. So any solution with  $\sigma \neq \tau$  will have to come in a one-parameter family, and hence it is impossible to have an isolated solution. This clashes with the fact that the index is 0, and means that generically there are no flows from  $\mathcal{A}_\sigma$  to  $\mathcal{A}_\tau$ .

We know an independent reason for this: the flow equation has a conserved quantity, namely the imaginary part of the holomorphic function  $\mathcal{I} = \lambda W(\mathcal{A})$ . For generic  $\lambda$ , conservation of  $\text{Im } \mathcal{I}$  makes a flow from  $\mathcal{A}_\sigma$  to  $\mathcal{A}_\tau$  impossible. The condition for a Stokes ray is precisely that a flow from  $\mathcal{A}_\sigma$  to  $\mathcal{A}_\tau$  is consistent with the conservation law. Taking  $\lambda = 2\pi \exp(-i\alpha)$ , a Stokes ray is characterized by some condition  $\alpha = \alpha_0$ . Instead of regarding  $\alpha$  as a constant in our flow equations (4.8) and (4.10), regard it as a variable that takes values in a small interval near  $\alpha = \alpha_0$ . This adds one variable to the problem without adding any new equations, and the index of the linearized problem is now 1. Thus the moduli space of solutions should generically be one-dimensional – corresponding to families of solutions, obtained from each other by the action of  $\mathbf{T}$ , that appear at  $\alpha = \alpha_0$ . To get back to a problem with zero index, we can identify two field configurations that differ by the action of  $\mathbf{T}$ . This removes one variable, and the number of solutions is now generically finite. Weighting the solutions by the sign of the determinant of the operator  $\tilde{L}$  that appears in the linearization around a solution, we arrive at the desired invariant  $\mathbf{m}_{\sigma\tau}$ .

Concretely, adding one variable and removing another and replacing  $L$  by  $\tilde{L}$  has the following effect. Suppose we set  $\alpha = \alpha_0$  and find a solution of the flow equations. The linearization  $L$  around a solution of the flow equations will have a zero mode  $\psi$ , generated by the action of  $\mathbf{T}$  on the given solution. Hence  $\det L = 0$ . Now vary  $\alpha$  away from  $\alpha_0$ . The eigenvector  $\psi$  of  $L$  can be deformed with  $\alpha$  so as to remain an eigenvector, but its eigenvalue  $\lambda(\alpha)$  generically vanishes only at  $\alpha = \alpha_0$ . Replacing  $L$  by  $\tilde{L}$  means that in evaluating the determinant at  $\alpha = \alpha_0$ , the factor that comes from the eigenvector  $\psi$  is not  $\lambda(\alpha_0) = 0$  but  $(d\lambda/d\alpha)|_{\alpha=\alpha_0}$ , which is generically not zero. So  $\det \tilde{L}$  is generically nonzero and its sign is the contribution of a given  $\mathbf{T}$ -orbit of flow lines to  $\mathbf{m}_{\sigma\tau}$ .

We have defined an invariant  $\mathbf{n}_\sigma$  associated to the flow equations on a half-line, and another invariant  $\mathbf{m}_{\sigma\tau}$  associated to the same equations on the full real line. The relation between these invariants is that in crossing a Stokes ray at which there is a net number  $\mathbf{m}_{\sigma\tau}$  of downward flow lines from  $\mathcal{A}_\sigma$  to  $\mathcal{A}_\tau$ , the invariant  $\mathbf{n}_\tau$  jumps by  $\mathbf{n}_\tau \rightarrow \mathbf{n}_\tau - \mathbf{n}_\sigma \mathbf{m}_{\sigma\tau}$ . As explained in the discussion of eqn. (3.20), where  $\mathbf{m}_{+-}$  is taken to be  $\pm 1$ , this jumping of the coefficients  $\mathbf{n}_\tau$  compensates for the jumping in the cycles  $\mathcal{J}_\sigma$  and ensures that the sum  $\sum_\sigma \mathbf{n}_\sigma \mathcal{J}_\sigma$  is invariant. In infinite dimensions, it seems most concise to describe the jumping phenomenon in terms of the behavior of the coefficients.

Just as in Morse theory with finitely many variables, the jumping happens because at

$\alpha = \alpha_0$  the set of points that can be reached by downward flow from  $\mathcal{A}_\sigma$  ceases to be closed – its closure contains  $\mathcal{A}_\tau$ . A downward flow from  $\mathcal{C}$  to  $\mathcal{A}_\tau$  can split into a pair of flows, first from  $\mathcal{C}$  to  $\mathcal{A}_\sigma$  and then from  $\mathcal{A}_\sigma$  to  $\mathcal{A}_\tau$ . This leads to the “wall-crossing” behavior  $\mathbf{n}_\tau \rightarrow \mathbf{n}_\tau - \mathbf{n}_\sigma \mathbf{m}_{\sigma\tau}$ . See for example [41] in the context of flows with finitely many variables. Ellipticity means that such statements about Morse theory carry over to the infinite-dimensional case. This is an important fact in the context of Floer theory (of three-manifolds and of loop spaces of symplectic manifolds). See for example [46, 47] for reviews. Hopefully the usual proofs in the context of Floer theory can be adapted to the present situation.

## 4.2 Knots And Morse Theory

Beginning in section 2.5.3, we described the analytic continuation of Chern-Simons knot invariants in a rather informal way. To make this more precise, we need to incorporate in the language of oscillatory integrals the holonomy functions that are included in the path integral to define knot invariants. How to do this has been briefly explained in [51], section 3.3, and in [54], pp. 1218-9, and in much more detail in [34]. We consider first the case of a compact gauge group  $H$  with connection  $A$ .

For a knot  $K$  labeled by a representation  $R$ , the basic factor is

$$\mathrm{Tr}_R P \exp \left( - \oint_K A \right). \quad (4.15)$$

According to the Borel-Weil-Bott theorem, letting  $T$  denote a maximal torus in  $H$ , the representation  $R$  can be constructed as the space of holomorphic sections of a certain holomorphic line bundle  $\mathcal{L}_R$  over the flag manifold  $H/T$ . This means that there is a quantum mechanical problem with  $H$  symmetry for which the space of physical states is the representation  $R$ . Writing a path integral representation of this system and coupling it to the gauge field  $A$  will give the representation that we need of the holonomy function.

The line bundle  $\mathcal{L}_R \rightarrow H/T$  admits an essentially unique  $H$ -invariant connection that we will call  $a_R$ . We can describe as follows a classical problem with  $R$  symmetry whose quantization leads, in view of the Borel-Weil-Bott theorem, to the representation  $R$ . Let  $S$  be a one-manifold and let  $\rho : S \rightarrow H/T$  be a map. On the space of such maps, we define an action  $I_R(\rho) = \int_S \rho^*(a_R)$ . For  $S$  a circle, this is defined mod  $2\pi$ , so that  $\exp(iI_R(\rho))$  is well-defined. (If instead  $S$  has boundaries,  $\exp(iI_R(\rho))$  is not a complex number but a transition amplitude between quantum wavefunctions; we will not need the details here.) Since  $I_R(\rho)$  is invariant under reparametrizations of  $S$ , this system is a topological field theory in the one-dimensional sense; the Hamiltonian vanishes.



The basic path integral for this system is the integral over maps  $\rho : S \rightarrow H/T$ :

$$\int D\rho \exp(iI_R(\rho)). \quad (4.16)$$

For  $S$  a circle, the value of the path integral can be determined by observing that the path integral determines a trace in the physical Hilbert space. The physical Hilbert space is simply the representation  $R$ , and as the Hamiltonian vanishes, the trace is simply the trace of the identity operator. So the value of the path integral is  $d_R$ , the dimension of the representation  $R$ .

We want to generalize this to couple to a gauge field  $A$  and to compute the trace not of the identity operator but of the holonomy operator  $\text{Tr}_R P \exp(-\oint_S A)$ . To do this, we simply couple the gauge field  $A$  to the  $\rho$  field in a gauge-invariant fashion. Quantum mechanically, we would couple  $A$  to the matrices that generate the action of  $H$  in the representation  $R$ . The classical limit of these matrices is the moment function  $\mu_R$  associated to the action of  $H$  on  $H/T$ .  $\mu_R$  is a map from  $H/T$  to the dual of the Lie algebra  $\mathfrak{h}$ . So for  $h \in \mathfrak{h}$ , we have a real valued function  $\mu_R(h)$  on  $H/T$ . It obeys  $d\mu_R(h) = \iota_h da_R$  (where  $\iota_h$  is contraction with the vector field on  $H/T$  that corresponds to  $h$ ). As  $A$  is  $\mathfrak{h}$ -valued, there is a natural pairing  $(A, \mu_R)$ . The gauge-invariant extension of the action  $I_R(\rho)$  is  $I_R(\rho, A) = I_R(\rho) - \int_S (A, \mu_R)$ . (A more intrinsic description of the geometry involved in defining this action can be found in [54]; for more detail see also [34].) Thus we arrive at a path integral representation of the holonomy:

$$\text{Tr}_R P \exp\left(-\oint_S A\right) = \int D\rho \exp(iI_R(\rho, A)). \quad (4.17)$$

Now we can rewrite the basic formula (2.19) for the Chern-Simons invariant of a knot in a convenient form:

$$Z_H(K; k) = \int_{\mathcal{U}} DA \int D\rho \exp(ikW(A) + iI_R(\rho, A)). \quad (4.18)$$

What we have gained is that now, although we are integrating over a larger number of variables, the integrand is simpler: it is just the exponential of a local functional of the fields. In other words, we have reduced the problem to an oscillatory integral (with more variables), and we can apply the Morse theory machinery rather as if no knot were present.

As usual, the main step in analytic continuation is to extend the action to a holomorphic function of complex variables. We analytically continue the connection  $A$  to a complex-valued connection  $\mathcal{A}$  and we replace the flag manifold  $H/T$  with  $H_{\mathbb{C}}/T_{\mathbb{C}}$ , where  $T_{\mathbb{C}}$  is a complex maximal torus in  $H_{\mathbb{C}}$ . The analytically continued integral thus takes the form

$$Z_H(K; k) = \int_{\mathcal{C}} D\mathcal{A} \int D\rho \exp(ikW(\mathcal{A}) + iI_R(\rho, \mathcal{A})), \quad (4.19)$$

where  $\mathcal{C}$  is an integration cycle in the space of fields  $\mathcal{A}, \rho$ .

Any integration cycle  $\mathcal{C}$ , including the original real integration cycle of Chern-Simons theory in the presence of a knot, can be expressed in terms of Lefschetz thimbles in the usual way. Such cycles are convenient for several reasons. One reason, discussed in section 3.1.3, is that it is straightforward to determine the asymptotic value of the integral over such a cycle, in the semiclassical region in which the action is large. What is this region? In the absence of a knot, the action is proportional to  $k$ , so the semiclassical region is the region of large  $k$ . In the presence of a knot labeled by a representation  $R$  with highest weight  $\lambda_R$ , the action  $I_R(\rho, A)$  is proportional to  $\lambda_R$ , since it is linear in  $a_R$  and  $\mu_R$  which are both proportional to  $\lambda_R$ . So in the presence of a knot, the natural semiclassical limit is obtained by taking  $k$  and  $\lambda_R$  to infinity with a fixed ratio. The Morse theory machinery will enable us to understand the asymptotic behavior of the analytically continued Chern-Simons path integral in this limit.

#### 4.2.1 Classical Equations In The Presence Of A Knot

In the presence of a knot  $K$ , the classical equation for  $A$  receives an extra contribution and no longer asserts simply that  $F$  is flat. Rather, the equation receives an extra contribution that is supported on  $K$ :

$$\frac{kF}{2\pi} = \delta_K \mu_R. \quad (4.20)$$

(Here  $\delta_K$  is a delta function that is Poincaré dual to  $K$ , and we use the linear form  $\text{Tr}$  to identify the Lie algebra  $\mathfrak{h}$  with its dual; this is also done in eqn. (4.22) below.) The fact that the curvature is a delta function supported on  $K$  means that the connection is flat away from  $K$  and has a monodromy around  $K$ . Exponentiating the delta function in the curvature, this monodromy is

$$U = \exp(2\pi\mu_R/k). \quad (4.21)$$

To understand this result, recall that  $H/T$  is a homogeneous space, and that the conjugacy class of  $\mu_R$  (as an element of  $\mathfrak{h}^*$ ) is constant and depends only on  $R$ . So eqn. (4.21) really defines a conjugacy class in  $H$ ; the monodromy around the knot must lie in this conjugacy class. We note that the conjugacy class in question depends only on the ratio  $\mu_R/k$ .

Let us make this more explicit for  $H = SU(2)$ . Take  $R$  to be the  $n$ -dimensional irreducible representation of  $SU(2)$ . Its highest weight is

$$\frac{n-1}{2} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \quad (4.22)$$

(To keep the structure constants real, we represent a real Lie algebra such as  $\mathfrak{su}(2)$  by anti-hermitian matrices.) Up to conjugacy, this is  $\mu_R$  for the representation  $R$ . So the monodromy

around a knot labeled by this representation is

$$\exp\left(\frac{i\pi(n-1)}{k}\begin{pmatrix}1 & 0 \\ 0 & -1\end{pmatrix}\right). \quad (4.23)$$

We will now make a remark on quantum corrections to this classical formula. In physics in general, quantum corrections cannot be entirely understood by shifting the values of classical parameters, but many quantum results in Chern-Simons theory can be understood more readily from a semiclassical point of view if one shifts  $n-1$  to  $n$  and  $k$  to  $k+2$  in (4.23). (The shift in  $k$  was discussed in section 2.7.1, and the shift in  $n$  has a similar interpretation. For the analogous shifts for noncompact Lie groups, see [48].) For our purposes, it is not necessary to take these shifts into account explicitly. Our goal is to determine the correct integration cycle for the path integral in the limit of large  $k$  and  $n$  with fixed ratio  $n/k$ . The answer to this question is not affected by shifts  $n \rightarrow n+1$  or  $k \rightarrow k+2$ . Those shifts enter when, after having found the right integration cycle for the path integral, one carries out the integral including one-loop effects. This will generate the shifts in  $n$  and  $k$  in the usual way.

#### 4.2.2 More On The Classical Equations

There are some important subtleties concerning the gauge fields obeying (4.20), especially after we analytically continue to a complex gauge field  $\mathcal{A}$ . In that case, the equation becomes

$$\frac{k\mathcal{F}}{2\pi} = \delta_K \mu_R, \quad (4.24)$$

where  $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ . We want to supplement this equation with a condition for vanishing of the moment map,

$$d_A \star_M \phi = 0, \quad (4.25)$$

where  $\mathcal{A} = A + i\phi$ . To describe what sort of gauge fields obey this equation, it suffices to work in a two-plane normal to  $K$ . We take polar coordinates  $r, \theta$  on this two-plane, with  $K$  understood to lie at  $r = 0$ . To obey (4.24) and (4.25), we can take

$$\begin{aligned} A &= \alpha d\theta + \dots \\ \phi &= -\gamma d\theta + \dots, \end{aligned} \quad (4.26)$$

where  $\alpha$  and  $\gamma$  are commuting elements of the real Lie algebra  $\mathfrak{h}$  with  $\alpha - i\gamma = \mu_R/2k$ , and the ellipses refer to terms less singular than  $1/r$ . (The angular form  $d\theta$  has a  $1/r$  singularity at  $r = 0$ .) This type of singularity in gauge theory was studied in detail<sup>19</sup> in [55], starting

---

<sup>19</sup>In that context, there was another term in  $\phi$  of the form  $\beta dr/r$ ;  $\beta$  can be removed by a gauge transformation (which is singular at  $r = 0$ ), and tentatively it seems that in the present problem we can just as well take  $\beta = 0$ .

in section 2.1. The most natural way to proceed is to fix a gauge in which  $\mu_R$  lies in the Lie algebra of a specified complex maximal torus  $T_{\mathbb{C}}$ . This can always be done globally along  $K$ . Then we require  $\mathcal{A}$  to have a singularity of the sort described in (4.26), and we only allow gauge transformations that preserve the gauge condition – in other words, gauge transformations that are  $T_{\mathbb{C}}$ -valued when restricted to  $K$ .

One important point, discussed in [55], is that by a gauge transformation that is singular at  $r = 0$ , one can shift  $\alpha$  by a coroot of  $H$ . This means that the formalism is invariant under shifting  $\mu_R$  by  $k$  times a coroot. For  $H = SU(2)$ , this corresponds to the symmetry

$$n \rightarrow n + 2k, \quad (4.27)$$

which actually undergoes a quantum correction to  $n \rightarrow n + 2(k + 2)$ .

There is an important subtlety, which we will describe first for  $H = SU(2)$ . Suppose that we vary  $\mu_R$  and  $k$  so that  $\mu_R/k$  goes to zero, or more generally so that it approaches a cocharacter (which we then remove by a singular gauge transformation, as mentioned in the last paragraph). This means that  $\alpha$  and  $\gamma$  go to zero, and a naive look at (4.26) might make one believe that in the limit,  $A$  and  $\phi$  become nonsingular. The correct conclusion, however, because of the ellipses in (4.26), which represent terms less singular than  $1/r$ , is only that in such a case,  $A$  and  $\phi$  are less singular than  $1/r$ . The equations  $\mathcal{F} = d_A \phi = 0$  have a solution with a singularity at  $r = 0$  that is (slightly) milder than  $1/r$ , and this less singular solution may be the result if one varies a family of solutions with the  $1/r$  singularity (4.26) so that the coefficient of  $1/r$  approaches zero. The singular solution in question, which was essentially first studied by Nahm [56], takes the form

$$\begin{aligned} A &= \frac{t_1 d\theta}{\ln r} \\ \phi &= \frac{t_2 dr}{r \ln r} - \frac{t_3 d\theta}{\ln r}, \end{aligned} \quad (4.28)$$

where  $[t_1, t_2] = t_3$ , and cyclic permutations thereof. This condition means that the  $t_i$  define a homomorphism from  $\mathfrak{su}(2)$  to  $\mathfrak{h}$ . For  $H = SU(2)$ , there is up to conjugation only one non-trivial homomorphism and we pick this one. (For other groups, all possible homomorphisms can arise, in general.) For this solution, the monodromy of  $\mathcal{A}$  around the singularity at  $r = 0$  is not trivial but is in the conjugacy class that contains the non-diagonalizable, unipotent element

$$V = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (4.29)$$

This is demonstrated by directly showing that, with  $A, \phi$  given in (4.28), the holonomy of  $\mathcal{A} = A + i\phi$  along any path that loops around  $r = 0$  is in the stated conjugacy class. For example, if the path is the circle  $r = r_0$  for some  $r_0$ , the monodromy is  $\exp(2\pi(t_1 - it_3)/\ln r_0)$ , which is in the stated conjugacy class. (For more detail, see section 3.8 of [55].)

What is happening here is that in the complex Lie group  $G = H_{\mathbb{C}}$ , as  $\mu_R/k \rightarrow 0$ , the conjugacy class that contains the element  $U = \exp(2\pi\mu_R/k)$  does not approach the conjugacy class of the identity; rather it approaches the conjugacy class of the element  $V$ . This may become clear if one considers the family of  $SL(2, \mathbb{C})$  elements

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda^{-1} \end{pmatrix}, \quad (4.30)$$

which are conjugate to the diagonal matrix  $\text{diag}(\lambda, \lambda^{-1})$  for generic  $\lambda$ , but to  $V$  if  $\lambda = 1$ . So if we consider a family of connections  $\mathcal{A}$  representing Chern-Simons critical points, parametrized by  $\mu_R/k$ , with monodromy  $U = \exp(2\pi\mu_R/k)$ , then as  $\mu_R/k \rightarrow 0$ , the monodromy of  $\mathcal{A}$  around the knot might approach 1, but it also might approach a unipotent element conjugate to  $V$ . The latter behavior is generic, since the conjugacy class of  $V$  has dimension two, while that of the identity consists of a single point.

Modulo a singular gauge transformation that shifts  $\alpha$  to zero,<sup>20</sup> the behavior is exactly the same if  $\mu_R/k$  approaches a value for which  $U \rightarrow -1$ . Given a family of flat connections  $\mathcal{A}$  of monodromy  $U$ , consider a limit in which  $U \rightarrow -1$ . The monodromy of  $\mathcal{A}$  may approach  $-1$ , but instead it may approach

$$-V = -\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (4.31)$$

We will encounter both cases in section 5.

We have described this exceptional behavior for  $SL(2, \mathbb{C})$ , but it has a close analog for any complex Lie group  $G$ . As long as  $U = \exp(2\pi\mu_R/k)$  is regular (that is, at long as the subgroup of  $G$  that it commutes with is a copy of  $T_{\mathbb{C}}$ ), Chern-Simons critical points in the presence of a knot correspond to flat connections whose monodromy around the knot is conjugate to  $U$ . But in general, the monodromy may be in any conjugacy class whose closure contains  $U$ . For more on this, see section 3.8 of [55].

### *The Flow Equations In The Presence Of A Knot*

Now let us briefly describe the flow equations in the presence of a knot. As usual, to formulate the flow equations, we introduce an extra “time” coordinate. So to study Chern-Simons theory on a three-manifold  $M$  in the presence of a knot  $K$ , we consider flow equations on the four-manifold  $X = \mathbb{R} \times M$ , with the knot now filling out the two-manifold  $\mathbb{R} \times K$ . The flow equations are the same as before, but now we must consider solutions with a

---

<sup>20</sup>Such a gauge transformation is single-valued in  $SO(3)$ , but not in  $SU(2)$ . It is not quite a symmetry of  $SU(2)$  Chern-Simons theory, but it is useful for analyzing the classical equations and describing critical points. For more on this, see section 4.2.7.

singularity along  $\mathbb{R} \times K$ . The singularity is given by (4.28) modulo terms less singular than  $1/r$ . In particular, for special values of  $\mu_R/k$ , we must allow a singularity that (perhaps after a singular gauge transformation to set  $\alpha$  to zero) takes the form (4.28). This seemingly exceptional case is important in applications, because the flat  $SL(2, \mathbb{C})$  connection associated to a hyperbolic metric on the knot complement has this sort of monodromy.

In effect, the flow equations have a codimension two singularity in four dimensions that can be modeled by the same surface operators studied in [55].

### 4.2.3 Analog For Noncompact Lie Groups

For the case of a compact Lie group  $H$ , we have described how to incorporate a knot labeled by a representation  $R$  in the framework of oscillatory integrals and Morse theory.

What happens if the gauge group is not compact, for example if it is  $G = H_{\mathbb{C}}$ ? If the representation  $R$  is unitary and arises by quantization of a coadjoint orbit  $\mathcal{O}$  of  $G$ , then there is a close analog of (4.17), with  $H/T$  replaced by  $\mathcal{O}$ . The resulting knot invariants, however, have not been seriously studied.

Let us consider instead the case that  $R$  is a finite-dimensional holomorphic representation of  $G$ , obtained by analytic continuation of an irreducible representation of  $H$ , as described in section 2.6.1. It is straightforward to bring such a representation into the Morse theory framework. We represent the holonomy around the knot in such a representation by the same factor as in (4.19), namely  $\int D\rho \exp(iI_R(\rho, \mathcal{A}))$ . So the path integral definition (2.29) of knot invariants associated to such a representation can be rewritten

$$Z_G(K; t, \tilde{t}) = \int_{\mathcal{C}} D\mathcal{A} D\tilde{\mathcal{A}} D\rho \exp\left(itW(\mathcal{A})/2 + i\tilde{t}W(\tilde{\mathcal{A}})/2\right) \exp(iI_R(\rho, \mathcal{A})). \quad (4.32)$$

Morse theory can now be used in the usual way to describe possible integration cycles for this integral. After analytic continuation,  $\mathcal{A}$  and  $\tilde{\mathcal{A}}$  are independent. Since  $I_R(\rho, \mathcal{A})$  is independent of  $\tilde{\mathcal{A}}$ , the presence of the knot does not affect the Euler-Lagrange equation for  $\tilde{\mathcal{A}}$ . The equation for a critical point thus tells us simply that  $\tilde{\mathcal{A}}$  is flat, while  $\mathcal{A}$  is a flat connection on the complement of  $K$  whose monodromy around  $K$  is in the conjugacy class of the element  $U = \exp(2\pi\mu_R/k)$  described in eqn. (4.21). (When  $U$  is not regular, this statement must be refined in a way described in section 4.2.2.) Of course, now we are supposed to take the conjugacy class of  $U$  in  $G$  rather than in  $H$ .

One point deserves note. As  $I_R(\rho, \mathcal{A})$  is not real, the exponential  $\exp(iI_R(\rho, \mathcal{A}))$  grows exponentially in some regions. As a result, even if  $t$  and  $\tilde{t}$  are complex conjugates, the path integral (4.32) does not converge if evaluated on the usual real cycle  $\tilde{\mathcal{A}} = \overline{\mathcal{A}}$ . Perhaps it

is possible to slightly displace the real cycle to make the integral converge, or perhaps the integral can only be defined using Lefschetz thimbles. As explained in section 2.6.1, in the case of knots in  $S^3$ , a suitable integration cycle can essentially be borrowed from the case of a compact gauge group.

A similar analysis applies if the gauge group is a noncompact real form  $H'$  of  $G = H_{\mathbb{C}}$ . If  $R$  is a unitary representation, one can imitate (4.17) using the appropriate coadjoint orbit. If  $R$  is a finite-dimensional representation of  $H'$  (obtained by analytic continuation from  $H$  to  $G = H_{\mathbb{C}}$  and then restriction to  $H'$ ), one can represent the knot by the now familiar factor  $\int D\rho \exp(iI_R(\rho, A'))$ , where now  $A'$  is  $\mathfrak{h}'$ -valued. Prior to analytic continuation,  $\rho$  is a map to  $H/T$ , which admits an action of  $G$  and hence of  $H'$ . After analytic continuation, the fields and the integrand of the path integral are the same as they are in the  $H$ -valued theory, but the integration cycle is different. This last statement holds exactly as it does in the absence of the knot.

#### 4.2.4 Classification Of Critical Points

To study the Chern-Simons path integral via the Morse theory machinery, a first step is to classify the critical points and evaluate the classical action at a critical point. As long as one considers only physical values of the coupling parameters, critical points correspond to flat connections modulo gauge transformation (perhaps with monodromy around a knot), and the Chern-Simons invariant  $W(A)$  of such a flat connection is understood as an element of  $\mathbb{R}/2\pi\mathbb{Z}$  (or  $\mathbb{C}/2\pi\mathbb{Z}$  in the case of a complex Lie group).

Upon analytic continuation away from physical values of the coupling, one must be more careful. Let us consider first the case that the gauge group is a compact Lie group  $H$  and one wishes to continue away from integer values of  $k$ . In this case, the integrand  $\exp(ikW(A))$  is invariant only under the subgroup of gauge transformations that leave  $W(A)$  fixed, rather than shifting it by a multiple of  $2\pi$ . If  $H$  is connected and simply-connected, this subgroup is the connected component of the group of gauge transformations. At any rate,  $\exp(ikW(A))$  should be understood as a function not on the space  $\mathcal{U}$  of gauge fields modulo gauge transformations, but on a cover  $\widehat{\mathcal{U}}$  on which the Chern-Simons invariant is well-defined. For connected and simply-connected  $H$ ,  $\widehat{\mathcal{U}}$  is the universal cover of  $\mathcal{U}$ .

So we need to consider critical points of the Chern-Simons function on  $\widehat{\mathcal{U}}$ . Such a critical point is simply a flat connection together with a lift of its Chern-Simons invariant to  $\mathbb{R}$  (or  $\mathbb{C}$ , if we consider complex critical points).

As an example, let us consider Chern-Simons theory on  $S^3$  in the presence of a knot  $K$ . Typically there are only finitely many critical orbits in the usual sense. In other words, up

to gauge transformation, there are typically only finitely many flat connections on the knot complement  $S^3 \setminus K$  with prescribed monodromy about  $K$ . However, when we analytically continue away from integer values of  $k$ , we need a real-valued Chern-Simons function, meaning that a critical point must be understood as a pair consisting of a flat connection and a lift of its Chern-Simons function to  $\mathbb{R}$ . This combined data determines a critical point of the function  $W(\mathcal{A})$  on  $\widehat{\mathcal{U}}$ . There always are infinitely many of these critical points, as the Chern-Simons invariant of any given flat connection can be lifted to  $\mathbb{R}$  in infinitely many ways. In section 5, we will study in detail the analytic continuation of the Chern-Simons path integral for some specific knots (the trefoil and the figure-eight knot) and we will see the important role played by distinct critical points that differ only in the value of their Chern-Simons invariant. This should come as no surprise, since in eqn. (3.73) we have already seen similar behavior in the analytic continuation of the Bessel function away from integer values of  $\kappa$ .

The same ideas hold if the gauge group is a non-compact real form  $H'$  of  $G = H_{\mathbb{C}}$ . However, the case that the gauge group is the complex Lie group  $G$  involves a new wrinkle. The Chern-Simons theory depends naturally on an integer-valued parameter  $\ell$  and a real parameter  $s$ , as shown in eqn. (2.2). As explained in section 2, to analytically continue to complex  $s$ , we view  $\tilde{\mathcal{A}}$ , which begins life as the complex conjugate of the gauge field  $\mathcal{A}$ , as an independent  $\mathfrak{g}$ -valued connection in its own right. The action is a linear combination of the two Chern-Simons actions  $W(\mathcal{A})$  and  $W(\tilde{\mathcal{A}})$ ,

$$I = \frac{t W(\mathcal{A})}{2} + \frac{\tilde{t} W(\tilde{\mathcal{A}})}{2}, \quad t = \ell + is, \quad \tilde{t} = \ell - is. \quad (4.33)$$

If we impose invariance under all gauge transformations, a critical point is a pair of flat connections  $\mathcal{A}, \tilde{\mathcal{A}}$ , up to gauge equivalence. In that case,  $W(\mathcal{A})$  and  $W(\tilde{\mathcal{A}})$  are each naturally defined as elements of  $\mathbb{C}/2\pi\mathbb{Z}$ . But then (for complex  $s$ ) the indeterminacy of  $I$  is not just an integer multiple of  $2\pi$  and the integrand  $\exp(iI)$  of the path integral is not defined.

One might think that to make sense of the path integral, we would have to lift both  $W(\mathcal{A})$  and  $W(\tilde{\mathcal{A}})$  from  $\mathbb{C}/2\pi\mathbb{Z}$  to  $\mathbb{C}$ . We indeed have to do this if the aim is to analytically continue to generic values of both  $\ell$  and  $s$ . However, there is a fairly natural problem, described in section 2, in which  $\ell$  remains an integer and only  $s$  is continued to complex values. In this case,  $(t + \tilde{t})/2$  remains an integer, so adding  $2\pi$  to both  $W(\mathcal{A})$  and  $W(\tilde{\mathcal{A}})$  causes no change in  $\exp(iI)$ . The most natural way to proceed is to regard  $\exp(iI)$  as a function on the smallest cover of  $\mathcal{U}_{\mathbb{C}} \times \mathcal{U}_{\mathbb{C}}$  where it can be defined. In practice, this means that a critical point is a pair of flat connections  $\mathcal{A}, \tilde{\mathcal{A}}$  together with a lift of the pair  $W(\mathcal{A}), W(\tilde{\mathcal{A}})$  to  $(\mathbb{C} \times \mathbb{C})/2\pi\mathbb{Z}$ , where  $\mathbb{Z}$  is diagonally embedded. In other words, we can shift both Chern-Simons invariants by a common multiple of  $2\pi$ , but we cannot shift them independently. We can use this freedom to arbitrarily specify how we want to lift the Chern-Simons invariant of either  $\mathcal{A}$  or  $\tilde{\mathcal{A}}$ , but then we must consider all lifts of the other.



It is instructive to consider the usual example of a knot  $K$  in  $S^3$ . In this case, as explained in section 4.2.3, a critical point in the usual sense corresponds to an ordinary flat connection  $\tilde{\mathcal{A}}$  on  $S^3$  plus a flat connection  $\mathcal{A}$  on  $S^3 \setminus K$  whose monodromy around  $K$  is in a prescribed conjugacy class. As  $S^3$  is simply connected,  $\tilde{\mathcal{A}}$  is gauge-equivalent to a trivial flat connection, and  $W(\tilde{\mathcal{A}})$  vanishes modulo  $2\pi$ . It is natural to fix the complex lift of  $W(\tilde{\mathcal{A}})$  to be zero. Once we do this, we have exhausted our freedom to shift the Chern-Simons invariants, and we must consider all possible complex lifts of  $W(\mathcal{A})$ .

The possible critical points, then, are classified by a flat connection  $\mathcal{A}$  on  $S^3 \setminus K$  with appropriate monodromy and with all possible choices for its Chern-Simons invariant. But this is the same set of critical points that we encounter in analytically continuing the theory on  $S^3$  with compact gauge group  $H$  in the presence of the same knot  $K$  labeled by the same representation  $R$ . What we have just described is a nonperturbative aspect of the relation between theories with gauge groups  $H$  or  $G = H_{\mathbb{C}}$  for the special case of knots in  $S^3$ . This relation was already discussed in section 2.6.

#### 4.2.5 Variation Of The Action

To use the Morse theory formalism, we need some information about the value of the Chern-Simons function  $W(\mathcal{A})$  for a classical solution. We take the gauge group to be a complex Lie group  $G = H_{\mathbb{C}}$ .

The following ideas prove to be useful. First of all, as sketched in fig. 10, the boundary of a small neighborhood of a knot  $K$  is a two-torus  $Y$ . On  $Y$ , we draw two circles – the meridian  $\widehat{\mathbf{m}}$ , which goes the short way around  $Y$  and has linking number 1 with  $K$ , and the longitude  $\widehat{\mathbf{l}}$ , which goes the long way around  $Y$  and which we choose to have linking number 0 with  $K$ .

Since the fundamental group of  $Y$  is abelian, if  $\mathcal{A}$  is a flat connection, its monodromies around  $\widehat{\mathbf{m}}$  and  $\widehat{\mathbf{l}}$  (which we will denote by the same symbols) are commuting elements of  $G$ . Generically, we can conjugate them to a common maximal torus  $T_{\mathbb{C}} \subset G$ . According to (4.21), for a knot labeled by a representation  $R$ ,  $\widehat{\mathbf{m}}$  is conjugate to  $U = \exp(2\pi\mu_R/k)$ . (More exactly, as explained in section 4.2.2, this is so if  $U$  is regular. In general,  $U$  is contained in the closure of the conjugacy class that contains  $\widehat{\mathbf{m}}$ .)

Let us consider in detail the case  $G = SL(2, \mathbb{C})$ . The maximal torus of  $G$  is a copy of  $\mathbb{C}^*$ , consisting of diagonal matrices

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}. \quad (4.34)$$

An important fact is that in a classical solution of Chern-Simons theory in the presence of a

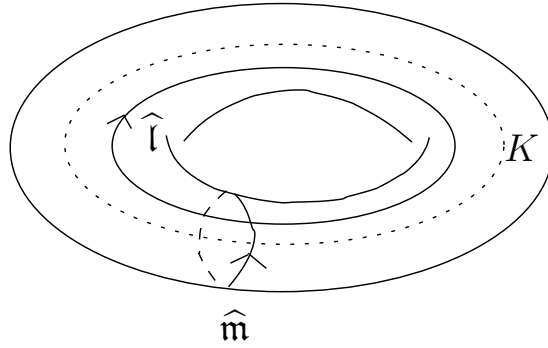


Figure 10: A knot  $K$  (dotted line) surrounded by a two-torus  $Y$ . On  $Y$  are drawn the meridian  $\widehat{\mathfrak{m}}$ , a circle whose linking number with  $K$  is 1, and the longitude  $\widehat{\mathfrak{l}}$ , whose linking number with  $K$  vanishes.

knot  $K$ , we have a little more structure than just a flat connection on the knot complement. In putting knots in the Morse theory framework, we introduced an  $H/T$  or  $G/T_{\mathbb{C}}$ -valued field  $\rho$  along  $K$ . In a classical solution,  $\rho$  is invariant under  $\widehat{\mathfrak{m}}$  and  $\widehat{\mathfrak{l}}$ , and once  $\widehat{\mathfrak{m}}$  and  $\widehat{\mathfrak{l}}$  are given, the choice of  $\rho$  amounts to a choice of one of the two eigenvalues of  $\widehat{\mathfrak{m}}$  and  $\widehat{\mathfrak{l}}$ , say the first eigenvalue (corresponding to  $\lambda$  as opposed to  $\lambda^{-1}$  in (4.34)). We write  $\mathfrak{m}$  and  $\mathfrak{l}$  for the eigenvalues of the loops  $\widehat{\mathfrak{m}}$  and  $\widehat{\mathfrak{l}}$  that correspond to  $\rho$ . According to (4.23), if  $R$  is the  $n$ -dimensional representation of  $SU(2)$ , then

$$\mathfrak{m} = \exp(i\pi(n-1)/k). \quad (4.35)$$

In section 5, we will use this formula in the semiclassical limit of large  $n, k$  with fixed ratio  $n/k$ . (In this limit, the difference between  $n$  and  $n-1$  is irrelevant.)

Once  $K$  is embedded in a three-manifold – which we will take to be simply a three-sphere –  $\mathfrak{l}$  and  $\mathfrak{m}$  are no longer independent. They are constrained by a requirement that an  $SL(2, \mathbb{C})$  flat connection on  $T$  with monodromies  $\widehat{\mathfrak{l}}$  and  $\widehat{\mathfrak{m}}$  should actually extend over the knot complement  $S^3 \setminus K$ .

One simple class of  $SL(2, \mathbb{C})$  flat connections are the abelian ones. These are the flat connections whose structure group reduces to  $\mathbb{C}^*$ , the group of diagonal matrices (4.34). They are classified very simply. Although the fundamental group of  $S^3 \setminus K$  can be very complicated, its first homology group is simply isomorphic to  $\mathbb{Z}$ , generated by  $\mathfrak{m}$ . This means that there is up to gauge transformation one flat  $\mathbb{C}^*$ -valued connection on  $S^3 \setminus K$  for every choice of  $\mathfrak{m}$ . Such a flat connection always has  $\mathfrak{l} = 1$ .

What about flat connections whose structure group does not reduce to  $\mathbb{C}^*$ ? Flat connec-

tions with triangular structure group

$$\begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix} \quad (4.36)$$

are unstable and need not be considered as critical points. What this really means is that to any strictly triangular flat connection there is a corresponding abelian one obtained by deleting the off-diagonal part of  $\mathcal{A}$ , and this should be considered instead. An important point which follows directly from the definition of the Chern-Simons function  $W(\mathcal{A})$  is that this function is invariant under this operation:

$$\begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix} \rightarrow \begin{pmatrix} \alpha & 0 \\ 0 & \gamma \end{pmatrix}. \quad (4.37)$$

There remain the irreducible  $SL(2, \mathbb{C})$  flat connections. (A flat connection is called irreducible if the only elements of  $SL(2, \mathbb{C})$  that commute with all monodromies are the central elements  $\pm 1$ .) An irreducible flat connection does not exist for all values of  $\mathfrak{l}, \mathfrak{m}$ . Rather, for each knot  $K$ , there is a Laurent polynomial  $\mathbf{A}(\mathfrak{l}, \mathfrak{m})$  (a polynomial in  $\mathfrak{l}, \mathfrak{l}^{-1}, \mathfrak{m}, \mathfrak{m}^{-1}$ ) such that the moduli space of irreducible  $SL(2, \mathbb{C})$  flat connections on  $S^3 \setminus K$  is described by an equation

$$\mathbf{A}(\mathfrak{l}, \mathfrak{m}) = 0 \quad (4.38)$$

which parametrizes values of  $\mathfrak{l}$  and  $\mathfrak{m}$  for which there exists such a connection.  $\mathbf{A}(\mathfrak{l}, \mathfrak{m})$ , which is often called the  $\mathbf{A}$  polynomial, is effectively computable for any given knot. We will make use of some examples in section 5. The  $\mathbf{A}$  polynomial is invariant under the exchange

$$\mathfrak{l} \rightarrow \mathfrak{l}^{-1}, \quad \mathfrak{m} \rightarrow \mathfrak{m}^{-1}, \quad (4.39)$$

which comes from a Weyl transformation exchanging the two eigenvalues in (4.34). Such a Weyl transformation does not leave invariant the field  $\rho$  that is defined along the knot, but the definition of the  $\mathbf{A}$  polynomial does not “know” about  $\rho$ . Another symmetry of the  $\mathbf{A}$  polynomial is

$$\mathfrak{l} \rightarrow \mathfrak{l}, \quad \mathfrak{m} \rightarrow -\mathfrak{m}. \quad (4.40)$$

This symmetry reflects the fact that on  $S^3 \setminus K$ , there exists a flat complex line bundle  $\mathcal{L}$  with monodromies  $\mathfrak{m} = -1, \mathfrak{l} = 1$ . If  $E \rightarrow S^3 \setminus K$  denotes a flat complex vector bundle of rank 2 and determinant 1, corresponding to a flat  $SL(2, \mathbb{C})$  connection, then the operation  $E \rightarrow E \otimes \mathcal{L}$  reverses the sign of  $\mathfrak{m}$  without changing  $\mathfrak{l}$ . It is not immediately apparent whether this operation leaves fixed the Chern-Simons function  $W(\mathcal{A})$  and therefore whether it is a symmetry of Chern-Simons theory. We return to this issue in section 4.2.7.

To describe the function  $W(\mathcal{A})$  once the  $\mathbf{A}$  polynomial is known (see [57] for a mathematical treatment), we simply use the variational principle for  $W(\mathcal{A})$ , which follows directly from its definition:

$$\delta W(\mathcal{A}) = \frac{1}{2\pi} \int_{S^3} \text{Tr } \delta \mathcal{A} \wedge \mathcal{F}. \quad (4.41)$$

On the other hand, for a flat connection on the knot complement, we have  $\mathcal{F} = (2\pi\mu_R/k)\delta_K = (\ln \widehat{\mathbf{m}})\delta_K$ . In the last step, we used the fact that  $\widehat{\mathbf{m}} = \exp(2\pi\mu_R/k)$ . So the right hand side of (4.41) reduces to  $\text{Tr} \oint_K \delta\mathcal{A} \ln \widehat{\mathbf{m}}$ . Here, once we diagonalize  $\mathbf{m}$ ,  $\mathcal{A}$  and  $\delta\mathcal{A}$  are also diagonal along  $K$ , and  $\oint_K \delta\mathcal{A} = \delta \ln \widehat{\mathbf{l}}$ . So the formula becomes  $\delta W = \text{Tr} (\delta \ln \widehat{\mathbf{l}}) \ln \widehat{\mathbf{m}}/2\pi$ . In expressing this result in terms of the eigenvalues  $\mathbf{l}$  and  $\mathbf{m}$  of the Lie algebra elements  $\widehat{\mathbf{l}}$  and  $\widehat{\mathbf{m}}$ , we get a factor of 2 from the trace:

$$\delta W(\mathcal{A}) = \frac{1}{\pi} \delta \ln \mathbf{l} \cdot \ln \mathbf{m}. \quad (4.42)$$

In practice, once the relation between  $\mathbf{l}$  and  $\mathbf{m}$  is known, the important results about  $W(\mathcal{A})$  can usually be obtained by integrating this formula. Some slightly subtle examples will be considered in section 5. Here we will practice with some easy cases.

For an abelian flat connection,  $\mathbf{l}$  is always 1, so  $0 = \delta \ln \mathbf{l} = \delta W(\mathcal{A})$ . So  $W(\mathcal{A})$  is constant as a function on the moduli space of abelian flat connections on  $S^3 \setminus K$ . It vanishes for a trivial flat connection ( $\mathbf{l} = \mathbf{m} = 1$ ) so it vanishes for any abelian flat connection. For another example, consider a unitary flat  $SL(2, \mathbb{C})$  connection – in other words, one whose holonomy lies in  $SU(2)$ . In this case,  $\mathbf{l}$  and  $\mathbf{m}$  are both of absolute value 1, so  $\ln \mathbf{m}$  and  $\ln \mathbf{l}$  are both imaginary. Hence,  $\delta W$  is real on this locus, consistent with the fact that  $W$  is real for an  $SU(2)$  flat connection.

Apart from the factor  $\exp(ikW(\mathcal{A}))$ , the integrand of the path integral in the presence of a knot also contains an additional factor  $\exp(iI_R(\rho, \mathcal{A}))$ , so we need to know how to evaluate this factor for a classical solution. In the usual gauge in which  $\rho$  is constant along  $K$ , this factor reduces to  $\exp(-i \int_K (\mu_R, \mathcal{A}))$ . This depends only on  $\mathbf{l}$  and can be understood as a character of the maximal torus in which  $\mathbf{l}$  takes values. Schematically we might write it as  $\mathbf{l}^{\mu_R}$ . For the  $n$ -dimensional representation of  $SU(2)$ , this factor is  $\mathbf{l}^{(n-1)}$ . Equivalently,

$$I_R(\rho, \mathcal{A}) = -i(n-1) \ln \mathbf{l}. \quad (4.43)$$

#### 4.2.6 Branches Of The Logarithm

An obvious point about this last formula is that  $\ln \mathbf{l}$  has many branches. This reflects the fact that the underlying action  $I_R(\rho, \mathcal{A})$  is similarly defined only modulo  $2\pi$ . There are two possible ways to proceed. If we wish to analytically continue to non-integral values of  $n$ , we should consider a critical point to be a pair consisting of a flat connection of suitable monodromy and a lift of  $\ln \mathbf{l}$  to  $\mathbb{C}$  (plus a similar lift of  $W(\mathcal{A})$  if we are analytically continuing in  $k$ ). This amounts to replacing the loop space of  $H/T$  or  $G/T_{\mathbb{C}}$  with its universal cover.

If we leave  $n$  as an integer and analytically continue only with respect to  $k$ , we need not consider the choice of a lift of  $\ln \mathbf{l}$  to be part of the definition of a critical point. Instead,

the multivaluedness of the logarithm enters in the following way. Consider the quantity  $\text{Im}(ikW(\mathcal{A}) + iI_R(\rho, \mathcal{A}))$  which one may expect to be conserved by the flow equations. The multivaluedness of the logarithm means that this quantity is actually only conserved modulo  $2\pi(n-1)\mathbb{Z}$ . It is conserved in any flow provided we make suitable choices for the branches of  $\ln \mathfrak{l}$ , but there may be no way to choose branches of  $\ln \mathfrak{l}$  that depend only on the choice of a critical point (and not on the choice of a flow line) such that  $\text{Im}(ikW(\mathcal{A}) + iI_R(\rho, \mathcal{A}))$  is conserved in every flow.

If instead we analytically continue with respect to  $n$ , we must consider the choice of branch of  $\ln \mathfrak{l}$  as part of the definition of a critical point and then the quantity  $\text{Im}(ikW(\mathcal{A}) + iI_R(\rho, \mathcal{A}))$  really is conserved by all flows.

See the end of section 5.2.4 for an example in which this issue becomes relevant.

#### 4.2.7 Tensoring With A Flat Line Bundle

As a further application of eqn. (4.42), we will return to the question of whether, for  $H = SU(2)$ , the operation  $E \rightarrow E \times \mathcal{L}$ , where  $\mathcal{L}$  is a flat line bundle of order 2, is a symmetry of Chern-Simons theory.

Let  $\Delta W(\mathcal{A})$  denote the change in  $W(\mathcal{A})$  under this operation. We can differentiate  $\Delta W(\mathcal{A})$  with respect to  $\mathcal{A}$  by making use of (4.42). The operation  $E \rightarrow E \otimes \mathcal{L}$  leaves  $\mathfrak{l}$  unchanged and (as it changes the sign of  $\mathfrak{m}$ ) shifts  $\ln \mathfrak{m}$  to  $\ln \mathfrak{m} - i\pi$ . So  $\delta \Delta W(\mathcal{A}) = (\delta \ln l / \pi)(-i\pi) = -i\delta \ln \mathfrak{l}$ . Integrating, we have  $\Delta W(\mathcal{A}) = -i \ln \mathfrak{l}$  up to an additive constant. In Chern-Simons theory on a general three-manifold (such as  $\mathbb{RP}^3$ ) the constant may be nonzero.<sup>21</sup> But it vanishes for the special case of a knot  $K \subset S^3$ . This can be proved by considering the special case that  $\mathcal{A}$  is a flat abelian connection (embedded in  $SL(2, \mathbb{C})$ ) with appropriate monodromy around  $K$ . Tensoring with  $\mathcal{L}$  gives another flat abelian connection  $\mathcal{A}'$ . The monodromies of  $\mathcal{A}$  and  $\mathcal{A}'$  around the longitude and the meridian are related by  $\mathfrak{m}' = -\mathfrak{m}$ ,  $\mathfrak{l}' = \mathfrak{l} = 1$ , and moreover  $W(\mathcal{A}) = W(\mathcal{A}') = 0$ . So for a knot in  $S^3$ , we have simply  $\Delta W(\mathcal{A}) = -i \ln \mathfrak{l}$ , and, making use of (4.43), we get a symmetry of  $kW(\mathcal{A}) + I_R(\rho, \mathcal{A})$  if we combine  $E \rightarrow E \otimes \mathcal{L}$  with  $n \rightarrow n + k$ . Allowing for a one-loop quantum shift described in section 2.7.1, the symmetry is actually

$$E \rightarrow E \otimes \mathcal{L}, \quad n \rightarrow n + k + 2. \quad (4.44)$$

This is actually the one place in the present paper that it is important to distinguish a knot (an embedded circle in  $S^3$ ) from a link (a disjoint union of embedded circles). The

---

<sup>21</sup>It is always an integer multiple of  $\pi$ , though not necessarily of  $2\pi$ . This can be proved by considering the Chern-Simons invariant for  $SO(3, \mathbb{C}) = SL(2, \mathbb{C})/\mathbb{Z}_2$ .

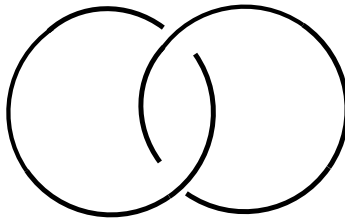


Figure 11: Two linked circles in  $S^3$ .

operation  $E \rightarrow E \otimes \mathcal{L}$ , where  $\mathcal{L}$  has monodromy around a knot, does not affect the longitude for that knot. But in the case of two or more disjoint embedded circles with nonzero linking (fig. 11), if  $\mathcal{L}$  has monodromy around one circle, then tensoring with  $\mathcal{L}$  may reverse the sign of the longitude for a second circle. So (4.44) is a symmetry of Chern-Simons theory for a knot, but not necessarily for a link.

The square of the operation (4.44)

$$n \rightarrow n + 2(k + 2) \tag{4.45}$$

is a universal symmetry of Chern-Simons theory (already mentioned in eqn. (4.27)). Indeed,  $\mathcal{L}^2$  is completely trivial, and tensoring with  $\mathcal{L}^2$  leaves  $\mathfrak{m}$  fixed.

## 5 Analytic Continuation Of The Colored Jones Polynomial

Here we will finally apply all this machinery to some specific examples. We will consider  $SU(2)$  Chern-Simons theory at level  $k$  on  $S^3$ , with a knot  $K$  labeled by the  $n$ -dimensional representation of  $SU(2)$ . We denote as  $J_n(q)$ , where  $q = \exp(2\pi i/(k+2))$ , the (unnormalized) Chern-Simons path integral in the presence of this knot.  $J_n(q)$  is essentially the colored Jones polynomial, though our normalization is not the most common one in the mathematical literature. (As explained at the end of section 2.5.3, the two normalizations differ by whether one chooses to divide by the Chern-Simons path integral for the unknot.)

We consider in detail two illustrative and relatively simple knots – the trefoil knot and the figure-eight knot (fig. 12). For background on these examples, see [17]. The trefoil knot is a basic example of a larger class of knots known as torus knots. They can be understood as orbits of a  $U(1)$  action on  $S^3$ , and this leads to exceptionally simple formulas for their quantum invariants, as has been explained recently via path integrals [34]. We expect similar behavior for other torus knots, though we consider only the trefoil in detail. However, torus

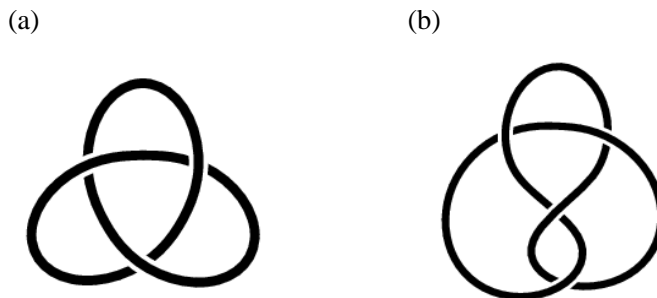


Figure 12: The trefoil knot (a) and the figure-eight knot (b). (From Wikipedia.)

knots are atypical. A more typical knot is one whose complement in  $S^3$  admits a complete hyperbolic metric.<sup>22</sup> The figure-eight knot (also called the  $4_1$  knot) is possibly the simplest example of a knot that is generic in this sense, and as a result its colored Jones polynomial has been studied extensively.

We will study the behavior of  $J_n(q)$  in the limit of large  $n, k$  with fixed ratio  $\gamma = n/k$ . We call this the semiclassical limit. For physical values of  $n$  and  $k$  – that is, positive integers – this limit can be analyzed by standard methods. One simply sums over contributions of real saddle points of the path integral; for physical values of  $n$  and  $k$ , the expansion around such real saddle points suffices for computing  $J_n(q)$  to all orders in an asymptotic expansion in  $1/k$  (or in  $1/n$  with  $\gamma$  fixed). The virtue of the Morse theory machinery of this paper is that it enables us to understand what happens in the semiclassical limit after analytic continuation away from integer values of  $k$  (and of  $n$ , if one so chooses).

## 5.1 Preliminaries

To apply the Morse theory machinery, we need to start at some point in the complex  $\gamma$  plane where we know how to express the integration cycle of Chern-Simons theory as a sum over Lefschetz thimbles. With this in hand, one can then vary  $\gamma$  and take Stokes phenomena into account.

If  $k$  and  $n$  are positive integers, so  $\gamma > 0$ , Chern-Simons theory is defined by an integral over  $SU(2)$ -valued connections modulo gauge transformations. In the framework of the present paper, this means that the integration cycle for  $\mathcal{A} = A + i\phi$  is the real cycle  $\phi = 0$ , which we call  $\mathcal{C}_{\mathbb{R}}$ . This cycle can be expressed in terms of Lefschetz thimbles as described

---

<sup>22</sup>The relevant metrics look near the knot like  $ds^2 = dz^2/z^2 + z^2(dx^2 + dy^2)$ , where the knot is located at  $z = 0$ , the “infrared” region in gauge/gravity duality. The “ultraviolet” region for  $z \rightarrow \infty$  is cut off by the embedding in  $S^3$ .

most precisely in eqn. (3.39):

$$\mathcal{C}_{\mathbb{R}} = \sum_{\sigma \in \Sigma_{\mathbb{R}}} \mathcal{J}_{\sigma} + \sum_{\sigma \in \Sigma_{-}} \mathbf{n}_{\sigma} \mathcal{J}_{\sigma}, \quad (5.1)$$

Here  $\Sigma$  is the set of critical points – or rather critical gauge orbits, but we will speak loosely of critical points. The critical points are flat connections on  $S^3 \setminus K$  with appropriate monodromy.  $\Sigma_{\mathbb{R}}$  consists of the real critical points (the ones with  $\phi = 0$ ), and  $\Sigma_{-}$  consists of complex critical points corresponding to flat connections  $\mathcal{A}$  for which the real part of the exponent of the path integral is negative. This exponent, including the contribution from the knot, is

$$\mathcal{I} = ikW(\mathcal{A}) + (n - 1) \ln \mathfrak{l}. \quad (5.2)$$

The coefficients  $\mathbf{n}_{\sigma}$  must be computed by counting solutions of flow equations. Because of the difficulty of actually doing this, our results for the figure-eight knot will be incomplete.

An important topological fact about  $SU(2)$  gauge theory is that for the case of a knot in  $S^3$  and monodromy sufficiently close to the identity, the only real critical point corresponds to the abelian flat connection  $\mathcal{A}_{\text{ab}}$ . So for sufficiently small values of  $\gamma = n/k$ , the contribution to  $\mathcal{C}_{\mathbb{R}}$  from the sum over  $\Sigma_{\mathbb{R}}$  is simply the cycle  $\mathcal{J}_{\text{ab}}$  associated to  $\mathcal{A}_{\text{ab}}$ .

$\mathcal{C}_{\mathbb{R}}$  is initially defined as a real integration cycle in  $\mathcal{U}_{\mathbb{C}}$ , the space of complex-valued connections. To analytically continue to complex values of  $k$ , we must lift and extend  $\mathcal{C}_{\mathbb{R}}$  to a cycle  $\mathcal{C} \subset \hat{\mathcal{U}}_{\mathbb{C}}$ ; recall that  $\hat{\mathcal{U}}_{\mathbb{C}}$  is a cover of  $\mathcal{U}_{\mathbb{C}}$  where the Chern-Simons function  $W(\mathcal{A})$  is well-defined as a complex-valued function. This process was described schematically in section 2.5. Concretely, analytic continuation can be made by choosing a lift to  $\mathbb{C}$  of the Chern-Simons invariant  $W(\mathcal{A}_{\sigma})$  for each critical point that appears in (5.1) with a nonzero coefficient. In section 2.5.3, a geometrical interpretation was given for a choice in which, for some given real number  $\alpha$ , one requires  $\alpha < \text{Im } \mathcal{I} < \alpha + 2\pi$  for all  $\sigma$ . The most general possible choice was described in section 2.5.2.

The choice that we will make of analytic continuation is guided by the fact that for fixed  $n$ , the colored Jones polynomial has a unique analytic continuation to a function  $J_n(q)$  with no essential singularity at  $k = \infty$  (or equivalently at  $q = 1$ ). Moreover [9], in the semiclassical limit,  $J_n(q)$  is bounded as  $n \rightarrow \infty$  for sufficiently small  $\gamma$ . We take these facts to mean that there is a unique choice of  $\mathcal{C}$  such that, in a sufficiently small neighborhood of  $\gamma = 0$ , the path integral has no exponentially growing contributions; it receives contributions only from critical points for which  $\text{Re } \mathcal{I}$  is zero or negative. This should hold regardless of the direction in which one approaches the point  $\gamma = 0$  in the complex plane.

The Chern-Simons invariant of the abelian flat connection  $\mathcal{A}_{\text{ab}}$  vanishes modulo  $2\pi$ . Avoiding exponential growth tells us how to lift  $\mathcal{J}_{\text{ab}}$  to  $\hat{\mathcal{U}}_{\mathbb{C}}$ : we must pick the lift such that the Chern-Simons invariant  $W(\mathcal{A}_{\text{ab}})$  is 0 (rather than a non-zero multiple of  $2\pi$ ). Similarly, we pick the branch of the logarithm such that  $\ln \mathfrak{l} = 0$ . Henceforth, when we refer to  $\mathcal{A}_{\text{ab}}$  as



a critical point in  $\widehat{\mathcal{U}}_{\mathbb{C}}$  or to  $\mathcal{J}_{\text{ab}}$  as a cycle in  $\widehat{\mathcal{U}}_{\mathbb{C}}$ , we have in mind this lifting, which we call the canonical lifting.

For any  $K$ , the contribution of  $\mathcal{A}_{\text{ab}}$  to the integration cycle is always precisely  $\mathcal{J}_{\text{ab}}$ , regardless of  $\gamma$ . In fact, Stokes phenomena never cause a jumping in the coefficients with which abelian critical points contribute to the integration cycle. This is explained in sections 3.3.4 and 3.4.2. Jumping definitely can occur for the coefficients of Lefschetz thimbles associated to nonabelian critical points.

For a typical knot, nonabelian critical points have nonzero  $\mathcal{I}$  near  $\gamma = 0$ ; let us assume this to be the case. At first sight, the argument of section 3.2.1 seems to show that if flat connections with nonzero  $\mathcal{I}$  contribute to the path integral near  $\gamma = 0$ , then an essential singularity for  $k \rightarrow \infty$  is unavoidable. If this were the case, we would conclude that in a neighborhood of  $\gamma = 0$ , the desired integration cycle  $\mathcal{C}$  must precisely coincide with  $\mathcal{J}_{\text{ab}}$  for all knots. There seem to be two fallacies in this reasoning.

First, the argument in section 3.2.1 assumed that the flow equations conserve  $\text{Im } \mathcal{I}$ . However, as discussed in section 4.2.6, in the presence of a knot, this is only so if one replaces the loop space of  $G/T_{\mathbb{C}}$  by its universal cover, as one should do if one wishes to analytically continue  $J_n(q)$  to non-integral values of  $n$ . In this case, a choice of branch of  $\ln \mathfrak{l}$  is regarded as part of the data specifying a critical point. If the concept of a critical point is understood in this way, then  $\text{Im } \mathcal{I}$  is conserved in the Chern-Simons flow equations and the argument of section 3.2.1 is applicable. Otherwise, we cannot make this argument.

At best (since we note another problem shortly), the conclusion from section 3.2.1 is that if  $J_n(q)$ , after being analytically continued to non-integral  $n$ , has no essential singularity for  $k \rightarrow \infty$  or  $q \rightarrow 1$ , then the integration cycle for sufficiently small  $\gamma$  is simply  $\mathcal{C} = \mathcal{J}_{\text{ab}}$ .

Unfortunately, although it seems clear in the framework of the present paper that  $J_n(q)$  can be analytically continued away from integer values of  $n$ , this does not seem to have been explored in the literature. The claim that  $J_n(q)$  has no essential singularity at  $k = \infty$  is only known in general for integer  $n$ , and we have little guidance about whether this remains so after analytic continuation in  $n$ .

The simplest example is the unknot  $K_0$ . The explicit formula<sup>23</sup> is

$$J_n(q; K_0) = \sqrt{\frac{2}{k+2}} \sin\left(\frac{\pi n}{k+2}\right). \quad (5.3)$$

Clearly, this function can be analytically continued in  $n$ , and the analytically continued

---

<sup>23</sup>Recall that in our normalization,  $J_n(q)$  is defined by a path integral rather than a ratio of path integrals; for an explanation of the relation of this to the usual normalization in the mathematical literature, see the last paragraph of section 2.5.3.

function has no essential singularity at  $k = \infty$ . For the trefoil, we will see at the end of section 5.2.4 that for  $n \notin \mathbb{Z}$ , there is an essential singularity at infinity,<sup>24</sup> but the situation for more general knots is very unclear.

The second problem is that Chern-Simons theory is more complicated than the situation considered in section 3.2.1. It is assumed there that the Stokes curves are straight rays, which is not the case in Chern-Simons theory, and perhaps more important, it is assumed that there are only finitely many critical points and Stokes curves. In Chern-Simons theory, once one analytically continues away from integer  $k$ , there are infinitely many critical orbits (corresponding to different liftings of the Chern-Simons invariant of a flat connection) meeting at  $\gamma = 0$ , and this can play an important role. In the case of the trefoil knot, we will see that along the positive imaginary  $\gamma$  axis, arbitrarily close to  $\gamma = 0$ , an infinite number of critical points must be included to describe the integration cycle.

We do not have a general theory of how to determine the integration cycle so as to avoid exponential growth of the path integral for  $k \rightarrow \infty$ , so we will have to take an exploratory point of view in our examples. In our study of the trefoil, we will show that  $\mathcal{J}_{\text{ab}}$  is indeed the correct integration cycle if one approaches  $\gamma = 0$  along any direction except the positive imaginary axis. For the figure-eight knot, we will not make an equally detailed analysis.

Once one knows the correct integration cycle near  $\gamma = 0$ , whether it is simply  $\mathcal{J}_{\text{ab}}$  or not, the appropriate integration cycle for any  $\gamma$  can be found in the usual way by varying  $\gamma$  and taking account of Stokes phenomena.

### 5.1.1 Local Models

Since studying the behavior near a singularity of the moduli space of flat connections is an important technique for understanding Stokes phenomena, it may be helpful for some readers if we briefly discuss the deformation theory of a flat bundle. (When we consider examples later, we do so in a more informal language, and some readers may omit the following discussion.)

Let  $E$  be a flat bundle over a manifold  $M$ , with structure group  $G$  and connection  $\mathcal{A} = \mathcal{A}_0$ . Let  $Y$  be the automorphism group of  $E$ , and set<sup>25</sup>  $V = H^1(M, \text{ad}(E))$ ,  $V' = H^2(M, \text{ad}(E))$ .  $V$  parametrizes first order deformations of  $E$ , and  $V'$  parametrizes obstructions to such deformations. In deformation theory, it is shown that there is a  $Y$ -invariant map  $\sigma : V \rightarrow V'$

---

<sup>24</sup>I am informed by C. Beasley that this can also be verified for more general torus knots using formulas of [34].

<sup>25</sup>Here  $\text{ad}(E)$  is the sheaf of infinitesimal gauge transformations, leaving fixed the structure along the knot if a knot is present.

such that the moduli space of deformations of  $E$  is  $\sigma^{-1}(0)/Y$ . (Rather than dividing by  $Y$ , it may be more convenient to think of the moduli space as the space  $\sigma^{-1}(0)$  with an action of  $Y$ .) We can expand  $\sigma$  as a power series in a variable  $v \in V$ . The lowest order term is quadratic and comes from the cup product. Higher order terms come from Massey products. (For example, see [58]. From a physical point of view, these higher order terms arise by integrating out massive modes of the gauge field; see section 4.2 of [59].)

If  $M$  is a three-manifold, then  $V'$  is dual to  $V$  and the discussion simplifies. In this case, there is a  $Y$ -invariant function  $\Phi$  on  $V$ , and the map  $\sigma$  is  $d\Phi$ , the differential of  $\Phi$ . Again  $\Phi$  can be expanded as a power series in  $v$ . The leading term is cubic and comes by evaluating the classical action  $\mathcal{I}(\mathcal{A})$  as a function on  $V$ . From a physical point of view,  $\Phi$  is the effective action for the zero modes of the gauge field; it contains cubic couplings that can be read off from  $\mathcal{I}(\mathcal{A})$ , and higher order couplings (described mathematically in terms of Massey products) that arise by integrating out massive modes.

In our application, there will be an additional parameter  $\gamma$ , controlling the monodromy around a knot. Suppose that the flat bundle  $E$  appears for  $\gamma = \gamma_0$ . To include  $\gamma$  in the above discussion, simply set  $\epsilon = \gamma - \gamma_0$ . The effective action  $\Phi$  can be extended from a function on  $V$  to a function on  $V \times \mathbb{C}_\epsilon$ , where  $\mathbb{C}_\epsilon$  is a copy of the complex plane parametrized by  $\epsilon$ , in such a way that flat connections, for any  $\epsilon$ , correspond to points in  $V \times \mathbb{C}_\epsilon$  at which the equation  $d_V \Phi = 0$  is obeyed. Here  $d_V$  is the exterior derivative in the  $V$  direction. In other words, flat connections for a given  $\epsilon$  correspond to critical points of the restriction of  $\Phi$  to the fiber of the projection  $V \times \mathbb{C}_\epsilon \rightarrow \mathbb{C}_\epsilon$  that lies above that value of  $\epsilon$ .

The question now arises of how many Massey products must be included to get a good model of the behavior near the flat bundle  $E$  that we started with. The answer, assuming<sup>26</sup> that for  $\epsilon = 0$  (and hence any  $\epsilon$  near 0), there are only finitely many flat bundles near  $E$ , is that one must go far enough to lift all degeneracies. If the cubic form obtained by evaluating  $W(\mathcal{A})$  as a function on  $V$  is nondegenerate, this gives a complete answer for  $\Phi$ ; otherwise, one must carry out a power series to higher order, including enough Massey products to get a nondegenerate result.

Singularity theory gives a convenient framework for this discussion. Pick a transversal  $\mathcal{U}$  to the gauge orbit of the flat bundle  $E$ . Consider  $\mathcal{I}(\mathcal{A})$  as a function on  $\mathcal{U}$ . This function has a singularity (that is, a degenerate critical point) at the point  $\mathcal{A} = \mathcal{A}_0$ . Though  $\mathcal{U}$  is infinite-dimensional, the singularity can be modeled in finite dimensions. In fact, after making some choices, one can embed  $V$  in  $\mathcal{U}$  and write  $\mathcal{U}$  as  $V \times N$ , where  $N$  is an infinite-dimensional space that parametrizes the nonzero modes of  $\mathcal{A}$ ; we assume that  $\mathcal{A} = \mathcal{A}_0$  corresponds to the product of the origin in  $V$  and a point in  $N$  that we call the origin. The decomposition

---

<sup>26</sup>This is not always true in three-manifold theory, since in general the moduli space of flat connections may have components of positive dimension. It is true in our applications involving knots. In general, one needs to include enough Massey products to lift whatever degeneracies do get lifted.

$\mathcal{U} = V \times N$  can be made in a  $Y$ -invariant fashion. Singularity theory says that one can choose the decomposition so that  $\mathcal{I}(\mathcal{A})$  near  $\mathcal{A} = \mathcal{A}_0$  is the sum of a function on  $V$  and a function on  $N$ , such that the function on  $V$  is the effective action  $\Phi$ , and the function on  $N$  has a nondegenerate critical point at the origin. The function  $\Phi : V \rightarrow \mathbb{C}$  thus gives a finite-dimensional model of the singularity of the function  $\mathcal{I}(\mathcal{A}) : \mathcal{U} \rightarrow \mathbb{C}$ . Because of the choices that are made in identifying  $\mathcal{U}$  as  $V \times N$ ,  $\Phi$  is only determined up to a diffeomorphism of  $V$  (more exactly, a  $Y$ -invariant diffeomorphism that leaves fixed the origin). Within its equivalence class, one can take  $\Phi$  to be a polynomial, determined by finitely many Massey products.

The importance of this for Chern-Simons theory is as follows. The integral over  $Y$  behaves in the semiclassical limit as an infinite-dimensional Gaussian integral (with perturbative corrections). It does not contribute to Stokes phenomena near  $\mathcal{A} = \mathcal{A}_0$ . The interesting subtleties in the behavior of the Chern-Simons path integral near the flat bundle  $E$  in the semiclassical limit are all contained in the integral over  $V$ .

We will encounter three examples of this construction. Conceptually, the most simple example arises for the figure-eight knot at  $\gamma = 1/3$ , where two nonabelian flat connections meet. The flat bundle  $E$  is irreducible, and its automorphism group  $Y$  is trivial. The space  $V$  is one-dimensional, and  $\Phi$  is the function  $\Phi = x^3/3 - \epsilon x$  related to the Airy integral. As  $\Phi$  is cubic, Massey products play no role. The other two examples involve the case that  $\mathcal{A}_0 = \mathcal{A}_{\text{ab}}$  is the abelian flat connection, so the group  $Y$  is  $\mathbb{C}^*$ . For either the trefoil or the figure-eight knot, there is a critical value of  $\gamma$  at which  $\mathcal{A}_{\text{ab}}$  meets one or more non-abelian flat connections. In each case  $V$  is two-dimensional and  $Y$  acts with weights 1 and  $-1$ . We write  $u, v$  for coordinates on  $V$  corresponding to these weights. For the trefoil knot, it turns out that the appropriate local model is  $\Phi = (uv)^2/2 - \epsilon uv$ . The quartic term in  $\Phi$  comes from a Massey product. For the figure-eight knot, one gets instead  $\Phi = (uv)^3/3 - \epsilon uv$ , where now a Massey product of rather high order is involved in generating the  $(uv)^3$  term.

### 5.1.2 Normalization Of The Path Integral

Instead of simply saying, as we have just done, that the contribution to the Chern-Simons path integral of a degenerate critical point can be modeled by a corresponding finite-dimensional integral, it is possible, as we will briefly sketch, to be more precise and to describe the normalization factors. The facts we will describe could be used to sharpen the comparisons that we will describe later between the Chern-Simons path integral and computations in the mathematical literature that are based on the exact solution of Chern-Simons theory, obtained by combinatorial methods. (We will not do so in this paper and the reader might choose to omit this explanation.)

We first recall the nondegenerate case. The contribution of a nondegenerate critical point to the Chern-Simons path integral was evaluated in [51], compared to the exact solution (and slightly corrected<sup>27</sup>) in [22], and further refined in [23, 27]. In Chern-Simons theory on a three-manifold  $M$  with compact gauge group  $H$ , let  $E \rightarrow M$  be an  $H$ -bundle and suppose that  $H^0(M, \text{ad}(E)) = H^1(M, \text{ad}(E)) = 0$ . The automorphism group  $P$  of  $E$  is then a finite group (in general, the Lie algebra of  $P$  is  $H^0(M, \text{ad}(E))$ ). Let  $\mathcal{I}_E = ikW$  be the action of Chern-Simons theory, evaluated for the critical point  $E$ . The contribution of the flat bundle  $E$  to the Chern-Simons path integral is then

$$\frac{1}{\#P} \exp(\mathcal{I}_E) \exp(i\pi\eta(\text{ad}(E))/4) \sqrt{T(\text{ad}(E))}. \quad (5.4)$$

where  $\eta(\text{ad}(E))$  is the Atiyah-Patodi-Singer  $\eta$ -invariant of the operator  $\star d + d\star$  acting on odd degree forms with values in  $\text{ad}(E)$ ;  $T(\text{ad}(E))$  is the Ray-Singer analytic torsion of this flat bundle; and  $\#P$  is the order of the finite group  $P$ . The expression (5.4) arises by performing the path integral for fluctuations around the flat bundle  $E$ , in the Gaussian approximation. Higher order corrections will multiply this expression by an asymptotic series in powers of  $1/k$  (the same is true for eqn. (5.5) below), as explored in [60, 61].

Now let us assume that  $E$  corresponds to a degenerate critical point. We continue to assume that  $H^0(M, \text{ad}(E)) = 0$ , but we take  $V = H^1(M, \text{ad}(E))$  to be nonzero. We can still carry out a Gaussian integral over the nonzero modes of the gauge field (the modes that in the terminology of section 5.1.1 parametrize  $N$ ). The result of integrating over the nonzero modes is still given by the formula (5.4), if properly interpreted. There is one essential point: for an acyclic flat connection, the torsion is a number, but in the general case, as explained in the original construction of analytic torsion [62], and used in the present context in [23], it is a linear form on a certain one-dimensional space – constructed as a tensor product of the determinants of the cohomology groups of the flat bundle in question. In the present situation, the only nonzero cohomology group is  $V = H^1(M, \text{ad}(E))$ , and  $\sqrt{T(\text{ad}(E))}$  is a linear form on  $\det V$ . It determines a translation-invariant measure on  $V$  that we will denote as  $dv$ .

What we want to integrate over  $V$ , using this measure, is essentially  $\exp(\Phi)$ , where the function  $\Phi$  was described in section 5.1.1. The generalization of (5.4) for  $V \neq 0$  is

$$\frac{1}{\#P} \exp(i\pi\eta(\text{ad}(E))/4) (k/2\pi^2)^{\dim V/2} \int_V dv \exp(\Phi), \quad (5.5)$$

where we replace  $\sqrt{T(\text{ad}(E))}$  by the measure  $dv$  and use it to integrate over  $V$ . (The factor  $\exp(\mathcal{I}_E)$  has disappeared as  $\mathcal{I}_E$  is simply the value of  $\Phi$  at  $v = 0$ . If  $v = 0$  were a nondegenerate critical point of  $\Phi$ , the factor  $k^{\dim V/2}$  would be canceled upon doing the Gaussian integral over  $V$ . That this factor is precisely a power of  $k/2\pi^2$  was found for

---

<sup>27</sup>A normalization was used in [51] in which  $\eta$  and  $\log T$  were both two times smaller than usual.

$H = SU(2)$  in [27]. This has apparently not been established in complete generality, though it can plausibly be done using localization formulas [32].) The idea of (5.5) is that the integrand  $\exp(i\pi\eta(\text{ad}(E))/2) dv \exp(\Phi)/\#P$  comes from the path integral over  $N$  and we complete the path integral by integrating over  $V$ .

In the general case (and in fact in two of the three examples mentioned at the end of section 5.1.1), we must consider a flat bundle  $E$  such that  $H^0(M, \text{ad}(E)) \neq 0$ . The automorphism group  $P$  of  $E$  then has dimension  $d_P = \dim H^0(M, \text{ad}(E))$ . Moreover,  $\sqrt{T(\text{ad}(E))}$  is now the ratio of a measure on  $V$  and one on the Lie algebra  $\mathfrak{p}$  of  $P$ . A convenient way to proceed is to pick a metric on  $\mathfrak{p}$  (such as the usual one in which short coroots have length squared two) and let  $\text{Vol}(P)$  be the volume of  $P$  using this metric. Once we have picked a metric and hence a measure on  $\mathfrak{p}$ ,  $\sqrt{T(\text{ad}(E))}$  can be regarded just as before as a measure  $dv$  on  $V$ . The generalization of (5.5) is then

$$\frac{1}{\text{Vol}(P)} \exp(i\pi\eta(\text{ad}(E))/4) (k/2\pi^2)^{(\dim V - \dim P)/2} \int_V dv \exp(\Phi). \quad (5.6)$$

The factor  $\exp(i\pi\eta(\text{ad}(E))/4)$  depends on the metric of  $M$ . The metric dependence can be removed (by subtracting a “counterterm”) once a framing (or two-framing) of  $M$  is picked; this was the approach in [51]. A canonical two-framing exists [63] and is often used. The formulas specialized to this two-framing were described in detail in [22] for  $H = SU(2)$ .

### 5.1.3 Orientation Of The Integration Cycle

As a last preliminary comment, we should perhaps discuss the fact that although the invariant  $J_n(q; K_0)$  of the unknot – given in eqn. (5.3) – has no essential singularity at  $k = \infty$ , it does have a square root singularity there. This square root singularity is common for all knots.<sup>28</sup> The meaning of the square root singularity is that under analytic continuation around  $k = \infty$ , the orientation of the integration cycle  $\mathcal{C}$  is reversed.

This reversal of orientation under analytic continuation is not an exotic phenomenon. It occurs for the simple Gaussian integral  $\int_{-\infty}^{\infty} dx \exp(ikx^2) = \sqrt{\pi/ik}$ , where the square root shows that under analytic continuation around  $k = \infty$ , the orientation of the integration cycle is reversed. (Concretely, this happens because if  $k$  is rotated by an angle  $\phi$  in the complex plane, the integration cycle in the  $x$ -plane should be rotated by an angle  $\phi/2$  to maintain convergence of the integral. For  $\phi = 2\pi$ , we get back the original integration cycle with opposite orientation.) The existence of this monodromy shows that the treatment of the orientation of the integration cycle will be subtle in Chern-Simons theory, but we will phrase our discussion to minimize the role of such details.

---

<sup>28</sup>It cancels in ratios  $J_n(q; K)/J_n(q; K_0)$ , which are the functions usually considered in the mathematical literature. These ratios are Laurent polynomials in  $q = \exp(2\pi i/(k+2))$ .

## 5.2 The Trefoil Knot

For both of our examples, the necessary facts about the representation of the fundamental group of the knot complement are described in [17]. We begin with the case that our knot  $K$  is the trefoil.

The fundamental group of  $S^3 \setminus K$  can be described by two generators  $\mathbf{x}, \mathbf{y}$  with one relation  $\mathbf{xyx} = \mathbf{yxy}$ .  $\mathbf{x}$  and  $\mathbf{y}$  are both conjugate to the meridian  $\widehat{\mathbf{m}}$ ; for further details, see fig. 11 in [17]. The longitude is  $\widehat{\mathbf{l}} = \mathbf{yx}^2\mathbf{yx}^{-4}$ .

For every choice of the conjugacy class of  $\widehat{\mathbf{m}}$  (apart from certain exceptional values that will be discussed), there are up to gauge transformation two  $SL(2, \mathbb{C})$  flat connections on  $S^3 \setminus K$ . The first is an abelian flat connection  $\mathcal{A}_{\text{ab}}$  with  $\mathbf{x} = \mathbf{y} = \text{diag}(\mathbf{m}, \mathbf{m}^{-1})$ . (As usual  $\mathbf{m}$  is the eigenvalue of the meridian  $\widehat{\mathbf{m}}$ .) Its Chern-Simons invariant vanishes.

There is also a nonabelian flat connection  $\mathcal{A}_{\text{nab}}$  whose monodromies are given up to conjugation by<sup>29</sup>

$$\mathbf{x} = \begin{pmatrix} \mathbf{m} & 1 \\ 0 & \mathbf{m}^{-1} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{m} & 0 \\ 1 - \mathbf{m}^2 - \mathbf{m}^{-2} & \mathbf{m}^{-1} \end{pmatrix}. \quad (5.7)$$

The eigenvalue of the longitude is

$$\mathbf{l} = -\mathbf{m}^{-6}. \quad (5.8)$$

According to (4.35), in  $SU(2)$  Chern-Simons theory,  $\mathbf{m}$  is of modulus 1, so we will need to analyze this case carefully. If  $|\mathbf{m}| = 1$ , the flat connection  $\mathcal{A}_{\text{nab}}$  has a structure group that reduces to either  $SU(2)$  or  $SL(2, \mathbb{R})$ . This can be anticipated as follows. The complex conjugate of a representation of the fundamental group in  $SL(2, \mathbb{C})$  is still a representation of the fundamental group in  $SL(2, \mathbb{C})$ , in general with  $\mathbf{m} \rightarrow \overline{\mathbf{m}}$ . However, if  $|\mathbf{m}| = 1$ , the operation  $\mathbf{m} \rightarrow \overline{\mathbf{m}}$  is equivalent to  $\mathbf{m} \rightarrow \mathbf{m}^{-1}$  and does not affect the conjugacy class of the meridian, which is

$$\begin{pmatrix} \mathbf{m} & 0 \\ 0 & \mathbf{m}^{-1} \end{pmatrix}. \quad (5.9)$$

So for  $|\mathbf{m}| = 1$ , representations of the fundamental group in  $SL(2, \mathbb{C})$  occur in complex conjugate pairs or else are isomorphic to their own complex conjugates. A representation with the latter property has a structure group that reduces to a real form of  $SL(2, \mathbb{C})$  – either  $SU(2)$  or  $SL(2, \mathbb{R})$ .

For the trefoil knot, as there is only one nonabelian representation up to conjugacy, it must be equivalent to its own complex conjugate. Concretely, let us try to conjugate the

---

<sup>29</sup>Our  $\mathbf{m}$  is called  $m^{1/2}$  in [17].

representation into  $SU(2)$ . We set

$$C = \begin{pmatrix} 1 & -(\mathfrak{m} - \mathfrak{m}^{-1})^{-1} \\ 0 & 1 \end{pmatrix}, \quad (5.10)$$

so

$$C\mathbf{x}C^{-1} = \begin{pmatrix} \mathfrak{m} & 0 \\ 0 & \mathfrak{m}^{-1} \end{pmatrix} \quad (5.11)$$

is an element of  $SU(2)$ . Then

$$C\mathbf{y}C^{-1} = \begin{pmatrix} (\mathfrak{m} - \mathfrak{m}^3)^{-1} & (\mathfrak{m} - \mathfrak{m}^{-1})^{-2} \\ 1 - \mathfrak{m}^2 - \mathfrak{m}^{-2} & (\mathfrak{m}^{-1} - \mathfrak{m}^{-3})^{-1} \end{pmatrix}. \quad (5.12)$$

If  $\mathbf{x}$  and  $\mathbf{y}$  can be conjugated into  $SU(2)$ , this can be done in a way that makes  $\mathbf{x}$  diagonal. Having put  $\mathbf{x}$  in the form (5.11), the only further conjugation that leaves it diagonal is by a diagonal matrix  $\text{diag}(\lambda, \lambda^{-1})$ . Conjugation by a diagonal matrix can map a unimodular matrix of the form

$$\begin{pmatrix} a & b \\ c & \bar{a} \end{pmatrix}, \quad (5.13)$$

such as  $C\mathbf{y}C^{-1}$ , into  $SU(2)$  if and only if  $b\bar{c} < 0$ . For  $|\mathfrak{m}| = 1$ , this is equivalent to

$$1 > \mathfrak{m}^2 + \mathfrak{m}^{-2}, \quad (5.14)$$

and that therefore is the condition that the flat connection  $\mathcal{A}_{\text{nab}}$  has structure group  $SU(2)$ . If instead

$$1 < \mathfrak{m}^2 + \mathfrak{m}^{-2}, \quad (5.15)$$

the structure group of  $\mathcal{A}_{\text{nab}}$  reduces to  $SL(2, \mathbb{R})$ .

### 5.2.1 Singularities Of The Moduli Space Of Flat Connections

The remaining case  $1 = \mathfrak{m}^2 + \mathfrak{m}^{-2}$  requires further study. This case corresponds to a singularity of the moduli space of flat connections, where different branches meet, as we will now describe.

The condition  $1 = \mathfrak{m}^2 + \mathfrak{m}^{-2}$  is equivalent to

$$\ln \mathfrak{m} = \pm i\pi/6, \quad \text{mod } \pi i\mathbb{Z}. \quad (5.16)$$

In this case, the monodromies of eqn. (5.7) are strictly upper triangular, and the structure group of a flat connection  $\mathcal{A}_{\text{nab}}$  with those monodromies reduces to the triangular subgroup of  $SL(2, \mathbb{C})$ , and not to either  $SU(2)$  or  $SL(2, \mathbb{R})$ . In this exceptional case, there is also



another flat connection  $\mathcal{A}'_{\text{nab}}$  that is inequivalent to  $\mathcal{A}_{\text{nab}}$ ; its monodromies are the lower triangular cousins of eqn. (5.7):

$$\mathbf{x} = \begin{pmatrix} \mathbf{m} & 0 \\ 1 & \mathbf{m}^{-1} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{m} & 0 \\ 0 & \mathbf{m}^{-1} \end{pmatrix}. \quad (5.17)$$

$\mathcal{A}'_{\text{nab}}$  is actually gauge equivalent to the complex conjugate of  $\mathcal{A}_{\text{nab}}$ ; indeed, eqn. (5.7), specialized to  $\mathbf{m}^2 + \mathbf{m}^{-2} = 1$ , can be transformed to (5.17) via complex conjugation composed with conjugation by the matrix  $\Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ .

Both  $\mathcal{A}_{\text{nab}}$  nor  $\mathcal{A}'_{\text{nab}}$  represent unstable critical points in Chern-Simons theory, as their structure groups are strictly triangular. In fact, when  $1 = \mathbf{m}^2 + \mathbf{m}^{-2}$ , the only semistable critical point is the abelian one  $\mathcal{A}_{\text{ab}}$  with  $\mathbf{x} = \mathbf{y} = \text{diag}(\mathbf{m}, \mathbf{m}^{-1})$ . Let us discuss the behavior of Chern-Simons theory in the vicinity of  $\mathcal{A}_{\text{ab}}$ . We think of  $\mathcal{A}_{\text{ab}}$  as a diagonal flat connection with  $\mathbb{C}^*$  structure group

$$\begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \quad (5.18)$$

embedded in  $SL(2, \mathbb{C})$ . If  $\mathbf{m}$  is fixed (by specifying the values of  $n$  and  $k$  in Chern-Simons gauge theory), then  $\mathcal{A}_{\text{ab}}$  has no diagonal deformations as a flat connection. However, it does have off-diagonal deformations. There is a one-dimensional space of deformations of upper triangular form

$$\begin{pmatrix} 0 & u \\ 0 & 0 \end{pmatrix}. \quad (5.19)$$

A deformation of  $\mathcal{A}_{\text{ab}}$  with  $u \neq 0$  gives, up to gauge transformation, the triangular flat connection  $\mathcal{A}_{\text{nab}}$ . The value of  $u$  does not matter, as long as it is nonzero, because  $\mathcal{A}_{\text{ab}}$  has a  $\mathbb{C}^*$  group of symmetries consisting of constant gauge transformations  $\text{diag}(\lambda, \lambda^{-1})$ , and these act on  $u$  by  $u \rightarrow \lambda^2 u$ . Similarly (or by complex conjugation composed with the gauge transformation  $\Omega$ ),  $\mathcal{A}_{\text{ab}}$  has a one-dimensional space of lower triangular deformations

$$\begin{pmatrix} 0 & 0 \\ -v & 0 \end{pmatrix}. \quad (5.20)$$

$v$  transforms under  $\mathbb{C}^*$  by  $v \rightarrow \lambda^{-2}v$ , so if  $v$  is nonzero, its value does not matter. A lower triangular deformation of  $\mathcal{A}_{\text{ab}}$  with nonzero  $v$  gives  $\mathcal{A}'_{\text{nab}}$ .

If we take  $u$  and  $v$  to be both nonzero, then only their product  $uv$  is  $\mathbb{C}^*$ -invariant. It is not possible to take both  $u, v \neq 0$  without varying  $\mathbf{m}$  away from  $\mathbf{m}^2 + \mathbf{m}^{-2} = 1$ . This is evident from (5.7), according to which a two-dimensional representation of the fundamental group of  $S^3 \setminus K$  that is not triangular has  $\mathbf{m}^2 + \mathbf{m}^{-2} \neq 1$ . Let  $\epsilon = 1 - \mathbf{m}^2 - \mathbf{m}^{-2}$ ;  $\epsilon$  is real if  $|\mathbf{m}| = 1$ . In the space of possible deformations of  $\mathcal{A}_{\text{ab}}$  to an irreducible flat connection, we can normalize the coordinates  $u, v$  so that the relation between  $u, v$ , and  $\epsilon$  is  $uv = \epsilon$ . This statement simply says that for a non-abelian flat connection  $\mathcal{A}_{\text{nab}}$ , the  $\mathbb{C}^*$ -invariant  $uv$  has a

unique value for all  $\epsilon$  (since  $\mathcal{A}_{\text{nab}}$  is unique up to gauge transformation), with  $uv = 0$  at  $\epsilon = 0$  (as we know from eqn. (5.7)), and finally the zero of  $uv$  at  $\epsilon = 0$  is simple (the monodromies (5.7) show that it is possible to have  $u \sim 1$ ,  $v \sim \epsilon$ , and so  $uv \sim \epsilon$ ; this should be no surprise as there is no symmetry that would force  $uv$  to have a zero of higher order).

Now let us specialize this to  $SU(2)$  Chern-Simons theory. In this case, the gauge field  $\mathcal{A}$  should be anti-hermitian, so we want  $v = \bar{u}$ . Also, the symmetry group of the abelian flat connection reduces to  $U(1) = SO(2)$ . If we set  $u = x + iy$ ,  $v = x - iy$ , where  $x, y$  are real coordinates on the space of first order deformations of the flat  $SU(2)$  connection  $\mathcal{A}_{\text{ab}}$ , then the relation between  $x, y$  and  $\epsilon$  is

$$x^2 + y^2 = \epsilon. \quad (5.21)$$

This equation has real solutions only for  $\epsilon \geq 0$ ; this encodes the fact that  $\mathcal{A}_{\text{nab}}$  can be conjugated to  $SU(2)$  only under that condition. Let us consider the exponent  $\mathcal{I} = ikW(\mathcal{A}) + iI_R(\rho, \mathcal{A})$  of the path integral of Chern-Simons theory for small  $\epsilon$  and near the abelian flat connection  $\mathcal{A}_{\text{ab}}$ . It has critical orbits with  $x = y = 0$  and with  $x^2 + y^2 = \epsilon$ . Moreover,  $\mathcal{I}$  vanishes for the abelian flat connection  $\mathcal{A}_{\text{ab}}$ , corresponding to  $x = y = 0$ , and in general is of order  $n$  in the semiclassical region of large  $n$  with fixed  $\gamma = n/k$ . Finally it has one nonzero critical orbit at  $x^2 + y^2 = \epsilon$ . These facts suggest that  $-i\mathcal{I}/n$  can be modeled near  $x = y = \epsilon = 0$  by the simple  $SO(2)$ -invariant function

$$F(x, y) = c \left( \frac{1}{2}(x^2 + y^2)^2 - \epsilon(x^2 + y^2) \right) \quad (5.22)$$

with a real constant  $c$ . We will verify this shortly using (5.26), but for the moment let us discuss the implications.

Generically, in expanding around a flat connection, the Chern-Simons function is a non-degenerate Morse function in the directions transverse to the gauge orbits. When this is so, after gauge fixing, the expansion around a given critical point can be approximated for large  $n$  and  $k$  as an infinite-dimensional (oscillatory) Gaussian integral. Nondegeneracy is visible in (5.22) for  $\epsilon \neq 0$ . As long as  $\epsilon$  does not vanish,  $x = y = 0$  is a nondegenerate critical point of  $F(x, y)$ , and near the critical orbit  $x^2 + y^2 = \epsilon$ , the function  $F(x, y)$  is nondegenerate in the directions transverse to the  $SO(2)$  action. However, this nondegeneracy and the Gaussian nature of the Chern-Simons path integral break down at  $\epsilon = 0$ , where the two critical orbits meet.

The expansion of the Chern-Simons path integral around a critical point can then no longer be approximated by a Gaussian, but the Gaussian approximation fails only for the two variables  $x$  and  $y$ . To improve upon the Gaussian approximation, we need only a suitable approximation to the dependence of  $\mathcal{I}$  on  $x$  and  $y$  near the singularity. Assuming that (5.22) is an appropriate approximation, we can model the integral over the variables  $x$  and  $y$  by

$$\int dx dy \exp(inF(x, y)). \quad (5.23)$$

(For the constant multiplying this integral, see (5.5).) The integral (5.23) was studied in section 3.4.2 as a basic example of an oscillatory integral with symmetry. There were two main results in that discussion.

First, in eqns. (3.59) and (3.60), we analyzed the integral by a simple scaling argument. According to (3.59), the integral is  $n^{-1/2}$  times a transcendental function (essentially an error function) of  $n^{1/2}\epsilon$ . In the limit  $n \rightarrow \infty$ ,  $\epsilon \rightarrow 0$  with  $n^{1/2}\epsilon$  fixed, the colored Jones polynomial will equal  $n^{1/2}$  times this error function (times a constant described in (5.5)). The factor of  $n^{1/2}$  is most obvious in the special case  $\epsilon = 0$ : according to (3.60), compared to the behavior with  $\epsilon \neq 0$ , the large  $n$  behavior of the integral gains an extra factor of  $n^{1/2}$  at  $\epsilon = 0$ . We can immediately apply this to the invariant  $J_n(k)$  at the values of  $n/k$  (such as  $1/6$  or  $5/6$ ) that correspond to  $\epsilon = 0$  or  $\mathbf{m}^2 + \mathbf{m}^{-2} = 1$ . We expect  $J_n(k)$  to grow with a factor of  $n^{1/2}$  at  $\epsilon = 0$ , relative to its behavior for generic real  $\epsilon$ :

$$\frac{J_n|_{\epsilon=0}}{J_n|_{\epsilon \neq 0}} \sim n^{1/2}, \quad n \rightarrow \infty. \quad (5.24)$$

This is a known result (Theorem 1.2 of [16]). We have expressed the result in terms of a ratio of values of  $J_n$  because our normalization of  $J_n$  (as a path integral rather than a ratio of path integrals) is not the most common one in the mathematical literature. With the usual normalization, the colored Jones polynomial is of order 1 for  $\epsilon < 0$  and of order  $n^{1/2}$  at  $\epsilon = 0$ .

Second, in section 3.4.2 we made a direct analysis of the Stokes phenomena in passing through  $\epsilon = 0$  (or more generally in crossing Stokes lines that pass through  $\epsilon = 0$ ). These results only depend on the singularity at  $\epsilon = 0$ , so they carry over immediately to Chern-Simons theory. Thus, we can understand some Stokes phenomena in  $SU(2)$  Chern-Simons theory without having to solve four-dimensional flow equations. Details are discussed in section 5.2.3.

Finally let us return to (5.22), but now consider it as a model of the behavior of  $SL(2, \mathbb{C})$  gauge fields near the singularity at  $\epsilon = 0$ . In  $SL(2, \mathbb{C})$ ,  $x$  and  $y$  are complex and the symmetry group is  $\mathbb{C}^*$ . Unlike the case of  $SU(2)$ , where the critical orbit  $x^2 + y^2 = \epsilon$  becomes equivalent to  $x = y = 0$  if we set  $\epsilon = 0$ , in  $SL(2, \mathbb{C})$  the equation  $x^2 + y^2 = 0$  can be solved with  $x, y \neq 0$ . Modulo the  $\mathbb{C}^*$  symmetry (which acts by  $x \pm iy \rightarrow \lambda^{\pm 2}(x \pm iy)$ ), the locus with  $x^2 + y^2 = 0$  but  $(x, y) \neq (0, 0)$  consists of two points,  $(x, y) = (1, \pm i)$ . These two points correspond to the triangular flat connections  $\mathcal{A}_{\text{nab}}$ ,  $\mathcal{A}'_{\text{nab}}$ . It is also possible to specialize the analysis to  $SL(2, \mathbb{R})$  gauge fields – the same singularity will also play a role in  $SL(2, \mathbb{R})$  Chern-Simons theory – but we leave this for the reader.

### 5.2.2 The Chern-Simons Function

To understand possible Stokes phenomena and to determine the asymptotic behavior of the Chern-Simons path integral, we must evaluate the Chern-Simons function  $W(\mathcal{A})$  for the flat connection  $\mathcal{A}_{\text{nab}}$ . For this, we use (4.42), which along with (5.8) gives  $\delta W(\mathcal{A}_{\text{nab}}) = -(6/\pi)\delta \mathbf{m}/\mathbf{m} \cdot \ln \mathbf{m} = -(3/\pi)\delta \ln^2 \mathbf{m}$ . So  $W(\mathcal{A}_{\text{nab}}) = -3 \ln^2 \mathbf{m}/\pi + \text{constant}$ . To determine the constant, we observe that in general  $W(\mathcal{A})$  vanishes mod  $2\pi$  for a representation with triangular structure group, which we have in the present problem at the values of  $\ln \mathbf{m}$  given in (5.16). Thus

$$W(\mathcal{A}_{\text{nab}}) = -\frac{3}{\pi} \ln^2 \mathbf{m} - \frac{\pi}{12} - 2\pi s, \quad s \in \mathbb{Z}. \quad (5.25)$$

As we are planning to analytically continue in  $k$ , we have made explicit the possibility of adding to  $W(\mathcal{A})$  an integer multiple of  $2\pi$ . We can now evaluate  $\mathcal{I} = ikW(\mathcal{A}) + (n-1) \ln \mathfrak{l}$ , using  $\mathfrak{l} = -\mathbf{m}^{-6}$  and  $\ln \mathbf{m} = i\pi(n-1)/k$  to evaluate  $\ln \mathfrak{l}$ . In the semiclassical limit of large  $n, k$ , we ignore the difference between  $n-1$  and  $n$ . However, we do want to take into account the multivaluedness of  $\ln \mathfrak{l}$ , as explained at the end of section 4.2.5. We get

$$\mathcal{I}(\mathcal{A}_{\text{nab}}) = ik \left( \frac{3\pi n^2}{k^2} - \frac{\pi}{12} - 2\pi s \right) - in \left( 6\pi \frac{n}{k} + 2\pi(r-1/2) \right), \quad r, s \in \mathbb{Z}. \quad (5.26)$$

Alternatively, in terms of  $\gamma = n/k$ , we can write

$$\mathcal{I}(\mathcal{A}_{\text{nab}}) = -in(3\pi\gamma + (\pi/12 + 2\pi s)\gamma^{-1} + 2\pi(r-1/2)). \quad (5.27)$$

We can use these formulas to justify (5.22). The relation between  $\mathbf{m}$  and  $\gamma$  is  $\ln \mathbf{m} = i\pi\gamma$ , so  $\epsilon = 1 - \mathbf{m}^2 - \mathbf{m}^{-2}$  vanishes precisely if

$$\gamma = \pm(1/6 + p), \quad p \in \mathbb{Z}. \quad (5.28)$$

For any number  $y$  that is congruent to  $\pm 1/6 \bmod \mathbb{Z}$ , the nonabelian flat connection  $\mathcal{A}_{\text{nab}}$  with  $\gamma = y$  and

$$\begin{aligned} s &= \frac{3y^2}{2} - \frac{1}{24} \\ r &= \frac{1}{2} - 3y \end{aligned} \quad (5.29)$$

is equivalent to an abelian flat connection  $\mathcal{A}_{\text{ab}}$  with its canonical lift (zero Chern-Simons invariant and  $\ln \mathfrak{l} = 0$ ). The values of  $r$  and  $s$  were chosen to ensure vanishing of both contributions to  $\mathcal{I}(\mathcal{A}_{\text{nab}})$ , namely  $ikW(\mathcal{A}_{\text{nab}})$  and  $(n-1) \ln \mathfrak{l}$ . With these values of  $r$  and  $s$ , we find for any  $\gamma$

$$\mathcal{I}(\mathcal{A}_{\text{nab}}) = -3\pi in\gamma^{-1}(\gamma - y)^2. \quad (5.30)$$

So for  $\gamma = y + \epsilon$ ,  $\mathcal{I}$  is of order  $n\epsilon^2$ . The fact that the coefficient of  $\epsilon^2$  is nonzero is the content of (5.22). Any  $SO(2)$ -invariant function  $F(x, y, \epsilon)$  that vanishes at  $x = y = 0$ , has only one critical orbit with  $(x, y) \neq 0$  that approaches  $(x, y) = 0$  for  $\epsilon \rightarrow 0$ , and is of order  $\epsilon^2$  at this orbit, can be put in the form (5.22) modulo a change of variables (an  $SO(2)$ -invariant redefinition of  $x$  and  $y$  that is the identity near  $x = y = 0$ ). Possible higher order terms in  $F$  are irrelevant in the sense of singularity theory and do not affect the estimate (5.24) of the critical behavior of the Chern-Simons path integral.

### 5.2.3 The Integration Cycle And The Asymptotic Behavior Of The Path Integral

An obvious consequence of (5.26) is that for real  $k$  and  $n$ , the nonabelian critical points have  $\mathcal{I}$  imaginary. Since  $\text{Re } \mathcal{I}$  is strictly increasing or decreasing along a flow line, and vanishes on the real integration cycle  $\mathcal{C}_{\mathbb{R}}$ , this implies that for real  $n$  and  $k$ , there are no flows between  $\mathcal{C}_{\mathbb{R}}$  and a nonabelian critical point. Hence for  $\gamma$  real and small, the integration cycle  $\mathcal{C}_{\mathbb{R}}$  is simply the Lefschetz thimble  $\mathcal{J}_{\text{ab}}$  associated to the real critical point  $\mathcal{A}_{\text{ab}}$ .

Extending this reasoning slightly, for real  $n$  and  $k$ , all critical points, abelian or non-abelian, have  $\text{Re } \mathcal{I} = 0$ . So there are no non-trivial flows between them.

How then can anything nontrivial happen for real values of  $n, k$ ? Something nontrivial must happen because for suitable real  $\gamma$ , the nonabelian flat connection  $\mathcal{A}_{\text{nab}}$  is  $SU(2)$ -valued and hence must contribute to the  $SU(2)$  Chern-Simons path integral.

The answer to this question is that on the real  $\gamma$  axis, jumping of the Lefschetz thimbles  $\mathcal{J}_{\sigma}$ , or of the coefficients in the usual expansion  $\mathcal{C} = \sum_{\sigma} \mathbf{n}_{\sigma} \mathcal{J}_{\sigma}$ , occurs only at values of  $\gamma$  where two different flat connections become coincident. At these points, which were identified in eqn. (5.28), Chern-Simons theory has jumping phenomena that can be understood from a finite-dimensional model of the appropriate singularity, as described in section 5.2.1.

Here we will analyze the integration cycle of the Chern-Simons path integral, and its asymptotic behavior, for  $\gamma$  close to the real axis. For a full analysis of what happens throughout the complex  $\gamma$  plane, one must analyze the Stokes curves; we postpone this to section 5.2.4.

We want to express the integration cycle of the Chern-Simons path integral for the trefoil knot in terms of Lefschetz thimbles associated to critical points. To this aim, we let  $\mathcal{J}_{\text{ab}}$  be the Lefschetz cycle associated to the abelian critical point  $\mathcal{A}_{\text{ab}}$  with its canonical lift.  $\mathcal{J}_{\text{ab}}$  is the appropriate integration cycle for sufficiently small  $\gamma = n/k$ . It will become clear that cycles associated to other ways of lifting the Chern-Simons invariant  $W(\mathcal{A}_{\text{ab}})$  will not play

any role. (This is actually a general property for any knot.)

On the other hand, we do have to consider distinct cycles associated to different ways of lifting the Chern-Simons invariant  $W(\mathcal{A}_{\text{nab}})$  of the nonabelian critical point  $\mathcal{A}_{\text{nab}}$ , in other words different choices of the integer  $s$  in (5.25). For each number  $x$  that is congruent to  $\pm 1/6 \bmod \mathbb{Z}$ , we write  $\mathcal{J}_x$  for the Lefschetz thimble associated to  $\mathcal{A}_{\text{nab}}$ , with its Chern-Simons invariant determined by the choice  $s = 3x^2/2 - 1/24$  as in (5.29). These cycles differ only by the lifting of the Chern-Simons invariant, so they are equivalent when the Chern-Simons level  $k$  is an integer.

We fix the relative orientations of the cycles  $\mathcal{J}_x$  by requiring that when  $k$  is an integer, the Chern-Simons path integral over  $\mathcal{J}_x$  is independent of  $x$ . Momentarily, we will fix the overall orientations of the  $\mathcal{J}_x$  by making a choice for  $\mathcal{J}_{1/6}$ .

Near  $\gamma = 0$ , the appropriate integration cycle of Chern-Simons theory is simply  $\mathcal{J}_{\text{ab}}$ . Now we will explore what happens when  $\gamma$  departs from zero, but only near the real  $\gamma$  axis, so that we only need to worry about jumping due to singularities. As we increase  $\gamma$  along the positive real axis, nothing happens until we reach the value  $\gamma = 1/6$ . At this point, we meet a singularity of a type analyzed in sections 3.4.2 and 5.2.1, associated with the fact that  $\mathcal{A}_{\text{ab}}$  and  $\mathcal{A}_{\text{nab}}$  become stably equivalent<sup>30</sup> at  $\gamma = 1/6$ . The resulting Stokes phenomenon was analyzed in eqn. (3.61):  $\mathcal{J}_{\text{ab}}$  jumps by  $\pm \mathcal{J}_{1/6}$ , depending on orientation. ( $\mathcal{J}_{\text{ab}}$  and  $\mathcal{J}_{1/6}$  correspond to  $\mathcal{J}_p$  and  $\mathcal{J}_o$  in (3.61).) We pick the orientation so that for  $\gamma$  slightly greater than  $1/6$ , the integration cycle of Chern-Simons theory is  $\mathcal{J}_{\text{ab}} + \mathcal{J}_{1/6}$ .

There is a simple intuitive meaning to this result. In general, for integer values of  $n, k$ , the path integral of Chern-Simons theory is given asymptotically for large  $n$  and fixed  $n/k$  by a sum over contributions of all  $SU(2)$ -valued flat connections, all with weight 1. For  $1/6 < n/k < 5/6$ , there are two such connections, namely  $\mathcal{A}_{\text{ab}}$  and  $\mathcal{A}_{\text{nab}}$ , and both must contribute with weight 1 to the integration cycle. As long as  $k \in \mathbb{Z}$ , the different  $\mathcal{J}_x$  are equivalent, and it does not matter which of them we include. But to describe the analytic continuation away from integer  $k$  in the region just to the right of the Stokes curve that passes through  $\gamma = 1/6$ , the right integration cycle is  $\mathcal{J}_{\text{ab}} + \mathcal{J}_{1/6}$ .

Increasing  $\gamma$  further, nothing new happens until we reach  $\gamma = 5/6$ . In crossing  $\gamma = 5/6$ , nothing happens to  $\mathcal{J}_{1/6}$ , but  $\mathcal{J}_{\text{ab}}$  will jump again, picking up  $\pm \mathcal{J}_{5/6}$ . We claim that (with the orientation convention chosen above) a minus sign is the right choice here, and hence that the appropriate integration cycle just beyond  $\gamma = 5/6$  is  $\mathcal{J}_{\text{ab}} + \mathcal{J}_{1/6} - \mathcal{J}_{5/6}$ .

To explain why the sign is negative, consider again the behavior for integer  $n, k$  with

---

<sup>30</sup>This means the following.  $\mathcal{A}_{\text{nab}}$  is triangular at  $\gamma = 1/6$  and its diagonal part coincides with  $\mathcal{A}_{\text{ab}}$ , so the gauge orbit containing  $\mathcal{A}_{\text{nab}}$  has  $\mathcal{A}_{\text{ab}}$  in its closure. Hence, to construct a sensible (Hausdorff) moduli space of flat connections, one has to consider  $\mathcal{A}_{\text{nab}}$  and  $\mathcal{A}_{\text{ab}}$  to be equivalent.

$5/6 < n/k < 7/6$ . In this region, the only  $SU(2)$ -valued flat connection is the abelian one, and hence the large  $n$ , fixed  $n/k$  behavior of the path integral must be given by the contribution of  $\mathcal{A}_{\text{ab}}$  alone. The contribution of  $\mathcal{A}_{\text{nab}}$  must vanish for integer  $k$ . Since  $\mathcal{J}_{1/6}$  and  $\mathcal{J}_{5/6}$ , which arise from two different liftings of  $\mathcal{A}_{\text{nab}}$ , are equivalent for integer  $k$ , the minus sign in the integration cycle  $\mathcal{J}_{\text{ab}} + \mathcal{J}_{1/6} - \mathcal{J}_{5/6}$  ensures this.

Increasing  $\gamma$  further, nothing happens until we reach  $\gamma = 7/6$ . At this point,  $\mathcal{J}_{\text{ab}}$  jumps by  $\pm \mathcal{J}_{7/6}$ . In this part of the real  $\gamma$  axis, there is again an  $SU(2)$ -valued flat connection that must contribute with weight 1. So the sign is  $+$  and the integration cycle is  $\mathcal{J}_{\text{ab}} + \mathcal{J}_{1/6} - \mathcal{J}_{5/6} + \mathcal{J}_{7/6}$ . The next jump is at  $\gamma = 11/6$ , where we must add  $-\mathcal{J}_{11/6}$  to the integration cycle.

In general, if  $x > 0$  is a number of the form  $\pm(1/6 + p)$ ,  $p \in \mathbb{Z}$ , then just to the right of  $\gamma = x$ , the appropriate integration cycle is

$$\mathcal{J}_{\text{ab}} + \mathcal{J}_{1/6} - \mathcal{J}_{5/6} \pm \cdots \pm \mathcal{J}_x. \quad (5.31)$$

Similarly, for negative  $\gamma$ , the integration cycle is  $\mathcal{J}_{\text{ab}} + \mathcal{J}_{-1/6} - \mathcal{J}_{-7/6} \pm \cdots \pm \mathcal{J}_x$ , for the most negative relevant value of  $x$ .

Having found the right integration cycle, it is straightforward to determine the asymptotic behavior of the Chern-Simons path integral for  $n \rightarrow \infty$  with fixed  $\gamma$ . In general, a flat connection  $\mathcal{A}_\sigma$  with  $\mathbf{n}_\sigma \neq 0$  makes a contribution to the path integral which in the semiclassical limit is of order  $\exp(\mathcal{I}_\sigma)$ , where  $\mathcal{I}_\sigma$  is the value of the exponent of the path integral at  $\mathcal{A} = \mathcal{A}_\sigma$ . Unless  $\mathcal{I}_\sigma = 0$ , this contribution has an exponential dependence on  $n$  for large  $n$ , since  $\mathcal{I}_\sigma$  is of order  $n$  in the semiclassical limit. In general,  $\exp(\mathcal{I}_\sigma)$  multiplies an asymptotic series in (not necessarily integral) powers of  $n$ . In the context of an expansion around a complex critical point, this series has been studied in detail in [18]. The dominant contribution to the path integral, for a given  $\gamma$ , is the one that maximizes  $\text{Re } \mathcal{I}_\sigma$ , subject to the condition  $\mathbf{n}_\sigma \neq 0$ .

In the case of the trefoil knot, the abelian flat connection  $\mathcal{A}_{\text{ab}}$  has  $\mathcal{I}_{\text{ab}} = 0$ . Its contribution  $Z_{\text{ab}}$  to the path integral is bounded (in fact, of order  $1/\sqrt{n}$ ) for  $n \rightarrow \infty$  with fixed  $\gamma$ . On the other hand, a non-abelian connection  $\mathcal{A}_{\text{nab}}$ , lifted so that it meets the abelian one at  $\gamma = y$ , has  $\mathcal{I}_y = -3\pi i n \gamma^{-1}(\gamma - y)^2$ , according to (5.30). So, assuming the relevant coefficient – which we will call  $\mathbf{n}_y$  – is not zero, the asymptotic behavior of its contribution to the Chern-Simons path integral is

$$Z_y \sim \exp(-3\pi i n \gamma^{-1}(\gamma - y)^2). \quad (5.32)$$

These contributions are all oscillatory as long as  $\gamma$  is on the real axis. (On the real axis, they actually dominate  $Z_{\text{ab}}$  in absolute value by a factor of  $\sqrt{n}$  for large  $n$ .) As soon as  $\gamma$  departs from the real axis, the  $Z_y$  grow or decay exponentially for  $n \rightarrow \infty$ , depending on the sign of

the real part of the exponent. If  $\gamma = a + ib$  with  $a, b$  real, the real part of the exponent is

$$\operatorname{Re} \log Z_y = 3\pi n b \left( 1 - \frac{y^2}{a^2 + b^2} \right). \quad (5.33)$$

Near the real axis, and assuming  $|a| > |y|$  (as otherwise  $\mathbf{n}_y = 0$ ), this is positive if and only if  $b > 0$ . So near the real axis, the condition for exponential growth of  $J_n(q)$  is that  $|a| > 1/6$  (so that one of the  $\mathbf{n}_y$  is nonzero) and  $\operatorname{Im} \gamma > 0$ . In this region,  $\operatorname{Re} \log Z_y$  is maximal for  $y = \pm 1/6$ , so this contribution dominates  $J_n(q)$ . In section 5.2.4, we will analyze the Stokes curves and extend this analysis to the rest of the complex  $\gamma$  plane.

What we have just obtained (except so far restricted to the region near the real  $\gamma$  axis) is the standard picture of the colored Jones polynomial of the trefoil knot. For example, see eqn. (2.2) of [7] or Proposition 3.2 of [16], where this invariant, and its cousins for other torus knots, is written as the sum of contributions that correspond to our  $Z_{\text{ab}}$  and  $Z_y$  (with the same values of  $y$ ). The starting point was an integral representation for the colored Jones polynomial of a torus knot which originally was deduced using formulas that come from braid group representations, quantum groups, and conformal field theory [5, 30]. This integral representation has recently been obtained directly from the Chern-Simons path integral [34].

#### 5.2.4 Stokes Curves

By a Stokes curve, we mean a curve  $l$  in the complex  $\gamma$  plane on which distinct critical points have the same value of the quantity  $\operatorname{Im} \mathcal{I}$  that is conserved in the flow equation. There are many such curves. But we only really care about Stokes curves across which some jumping occurs. It turns out that there are far fewer of these.

Our basic strategy will be to show that every Stokes curve intersects the real  $\gamma$  axis. But we already know from section 5.2.3 what sort of jumping occurs along the real axis. It takes the form  $\mathcal{J}_{\text{ab}} \rightarrow \mathcal{J}_{\text{ab}} \pm \mathcal{J}_y$ , where  $\mathcal{J}_y$  is a cycle associated to one of the lifts of  $\mathcal{A}_{\text{nab}}$ . This will enable us to get a general description of the integration cycle. In fact, it turns out that the integration cycle always takes the form of (5.31), or its cousin with negative  $x$ , with some  $x$  that depends on  $\gamma$  in a way that will emerge.

A preliminary observation is that there are no flows between abelian critical points differing in the values of  $r, s$ , since all such critical points have  $\operatorname{Re} \mathcal{I} = 0$ , but  $\operatorname{Re} \mathcal{I}$  is never constant along a non-trivial flow. So in any flow between two critical points, at least one of them is nonabelian.

To show that all Stokes curves intersect the real  $\gamma$  axis, we use the explicit formula for  $\mathcal{I}$  and the fact that  $\operatorname{Im} \mathcal{I}$  is conserved by the flow equations. Equivalently, the conserved



quantity is  $\text{Im } \mathcal{I}/\pi n$ . For a lift of  $\mathcal{A}_{\text{nab}}$ , this can be conveniently expressed in terms of  $\gamma = n/k$ :

$$\text{Im } \mathcal{I}/\pi n = -\text{Re} \left( 3\gamma + \left( \frac{1}{12} + 2s \right) \gamma^{-1} + 2r - 1 \right). \quad (5.34)$$

If  $\gamma = a + ib$  with  $a, b$  real, then

$$\text{Im } \mathcal{I}/\pi n = -3a - \left( \frac{1}{12} + 2s \right) \frac{a}{a^2 + b^2} - 2r + 1. \quad (5.35)$$

Let us first consider Stokes curves at which there might be flows between nonabelian critical points with different values of  $r, s$ . The change in  $\text{Im } \mathcal{I}/\pi n$  in such a process is

$$\Delta(\text{Im } \mathcal{I}/\pi n) = -2 \left( \frac{a \Delta s}{a^2 + b^2} + \Delta r \right), \quad (5.36)$$

where  $\Delta s, \Delta r$  are the jumps in  $s$  and  $r$ . Setting this to zero, we get

$$a^2 + b^2 + \frac{a \Delta s}{\Delta r} = 0. \quad (5.37)$$

The case  $\Delta s = 0$  is not of interest, for then vanishing of (5.36) implies  $\Delta r = 0$ , in which case the initial and final critical points are the same and the flow is trivial. If  $\Delta s$  and  $\Delta r$  are both nonzero, the Stokes curve is a circle that intersects the real  $\gamma$  axis in two points; if  $\Delta r = 0$ , it becomes the line  $a = 0$ , intersecting the real  $\gamma$  axis at one point. Either way, at an intersection point of one of these Stokes curves with the real axis, two nonabelian critical points with different values of  $r, s$  remain different, and there are no flows between them, since in general there are no non-trivial flows for real  $\gamma$ .

In general, for two critical points  $p_\sigma, p_\tau$ , along a Stokes curve  $l$  characterized by  $\text{Im } \mathcal{I}_\sigma = \text{Im } \mathcal{I}_\tau$ , the integer  $\mathbf{m}_{\sigma\tau}$  that counts flows between  $p_\sigma$  and  $p_\tau$  is constant, except possibly at a point along  $l$  at which a third critical point  $p_\nu$  has  $\text{Im } \mathcal{I}_\nu = \text{Im } \mathcal{I}_\sigma = \text{Im } \mathcal{I}_\tau$ . Such a point is a triple intersection of Stokes curve (equality of any two of  $\text{Im } \mathcal{I}_\sigma, \text{Im } \mathcal{I}_\tau$ , and  $\text{Im } \mathcal{I}_\nu$  defines such a curve). At a triple intersection point there can be jumping  $\mathbf{m}_{\sigma\tau} \rightarrow \mathbf{m}_{\sigma\tau} + \mathbf{m}_{\sigma\nu} \mathbf{m}_{\nu\tau}$ . Eqn. (5.37) implies that there are no triple intersection points involving different lifts of  $\mathcal{A}_{\text{nab}}$ . (Two of the curves defined by eqn. (5.37) can intersect away from the real  $\gamma$  axis only if they coincide identically, in which case the effects they produce can again be analyzed from what happens on the real axis.) We need not worry about the case that  $p_\nu$  is a lift of  $\mathcal{A}_{\text{ab}}$ , since in this case  $\mathbf{m}_{\nu\tau} = 0$ , as we learned in section 3.3.4.

Hence the net number of flows between nonabelian critical points differing in the values of  $r$  and  $s$  is zero anywhere along one of these Stokes curves. So there are no Stokes phenomena involving flows between different lifts of  $\mathcal{A}_{\text{nab}}$ .

The same type of argument can be used to constrain Stokes phenomena involving a flow between the abelian critical point and a nonabelian one. If we pick the canonical lift of  $\mathcal{A}_{\text{ab}}$ , it has  $\mathcal{I} = 0$  and the change in  $\text{Im } \mathcal{I}/\pi n$  is given in (5.35). Setting the right hand side of that equation to zero, we get the equation for a Stokes curve:

$$b^2 = -\frac{a(3a^2 + (2r - 1)a + (2s + 1/12))}{3a + 2r - 1}. \quad (5.38)$$

If we pick a different lift of  $\mathcal{A}_{\text{ab}}$ , we get the same formula with shifted values of  $r, s$ . However, a byproduct of the following calculation will be to show that different lifts of  $\mathcal{A}_{\text{ab}}$  do not enter the integration cycle; indeed, all jumping phenomena take the form of (3.61) and do not shift the “abelian” part of the integration cycle.

Eqn. (5.38) describes many Stokes curves, labeled by the choice of  $r$  and  $s$  and a component of the part of the  $a$  axis on which the right hand side of (5.38) is nonnegative. Such a component is either a closed interval whose endpoints are zeroes of the numerator  $a(3a^2 + (2r - 1)a + (2s + 1/12))$  or a half-open interval whose endpoints are such a zero and the zero of the denominator. Each component intersects the  $a$  axis, either in two points (the first case) or one point (the second case).

We now make the familiar argument. Since every Stokes curve  $l$  intersects the  $a$  axis, the jumping that occurs in crossing  $l$  anywhere can be deduced from what happens in crossing  $l$  where it meets that axis. If  $l$  intersects the  $a$  axis at a value of  $a$  not of the form  $\pm(1/6 + p)$ , with  $p \in \mathbb{Z}$ , then  $l$  is irrelevant because no jumping can occur at such a value of  $a$ . Even if  $l$  intersects the  $a$  axis only at a point of the form  $\pm(1/6 + p)$ , something non-trivial can happen only if  $r$  and  $s$  are of the form given in (5.29). The reason for this last statement is that as there are no non-trivial flows for real  $\gamma$ , all that can happen in crossing  $l$  at a point  $y = \pm(1/6 + p)$  is that some lift of  $\mathcal{A}_{\text{nab}}$  may coincide with  $\mathcal{A}_{\text{ab}}$  with its canonical lift. For this to happen,  $r$  and  $s$  must be related to  $y$  by (5.29).

So the relevant Stokes curves are determined by their intersection point  $y = \pm(1/6 + p)$  with the real axis; the curve is described by (5.38) with  $r, s$  given in (5.29). In this special case, (5.38) simplifies to

$$b^2 = -\frac{a(a - y)^2}{a - 2y}. \quad (5.39)$$

We now meet a situation described at the end of section 3.4.2. Eqn. (5.39) gives the condition that a flow is possible between  $\mathcal{A}_{\text{ab}}$  (with its canonical lifting) and  $\mathcal{A}_{\text{nab}}$  (with its chosen lifting). The flow will go from  $\mathcal{A}_{\text{ab}}$  to  $\mathcal{A}_{\text{nab}}$  if  $\text{Re } \mathcal{I}(\mathcal{A}_{\text{nab}})$  is negative, and in the other direction if  $\text{Re } \mathcal{I}(\mathcal{A}_{\text{nab}})$  is positive. However, we also know from the analysis in section 3.4.2 of a singularity of the type that occurs at  $a = y$  (or from a more general argument in section 3.3.4) that no jumping comes from flows from  $\mathcal{A}_{\text{nab}}$  to  $\mathcal{A}_{\text{ab}}$ . So just as in fig. 5, we can throw

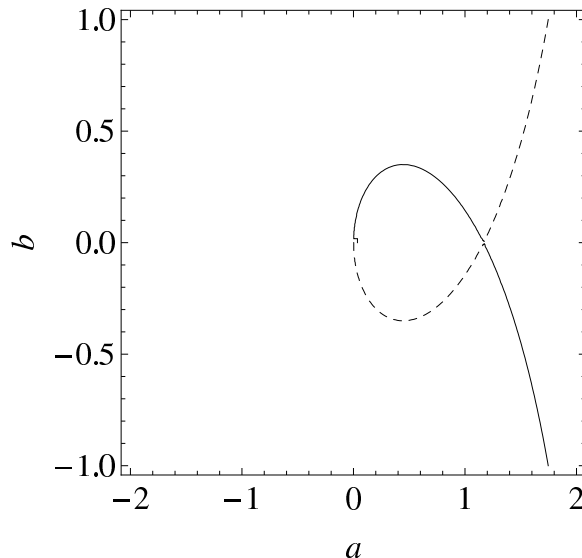


Figure 13: The Stokes curve (5.39) consists of both the solid and dashed curves, but just as in fig. 5, jumping occurs only across the solid part of the curve. (Only the solid parts, therefore, will be drawn in fig. 14.)

away the part of the curve (5.38) with  $\text{Re} \mathcal{I}(\mathcal{A}_{\text{nab}}) > 0$ . As sketched in fig. (13), the part of the curve that we keep is the part with  $yb(y - a) > 0$ .

Keeping only the “good” parts of these curves, they divide the complex  $\gamma$  plane in a way that is sketched in fig. 14. The important point here is that any point  $\gamma = \gamma_0$  that is not on one of the relevant Stokes curves can be connected to the  $a$  axis in a unique way (topologically) without crossing any of those curves. The integration cycle  $\mathcal{C}$  of Chern-Simons theory analytically continued to  $\gamma = \gamma_0$  is given by the same formula as on the part of the  $a$  axis to which  $\gamma_0$  can be connected without crossing a Stokes curve. This result for the integration cycle is in full accord with other analyses of the colored Jones polynomial of the trefoil knot. What we call the Stokes curve (5.39) – or more exactly the relevant portion of it with  $yb(y - a) > 0$  – is the curve across which an extra term appears in eqn. (2.2) of [7].

The analytically continued function  $J_n(q)$  grows exponentially in the semiclassical region whenever  $\text{Re} \mathcal{I} > 0$  for a critical point that contributes to the path integral. This happens precisely if  $\text{Im} \gamma > 0$  and  $|\gamma| > 1/6$ . (Those conditions are strong enough to ensure, using eqn. (5.39), that a critical point with  $y = 1/6$  or  $-1/6$  contributes to the path integral, and, using eqn. (5.33), that it has  $\text{Re} \mathcal{I} > 0$ .)

We are left with only one more issue, which is the behavior along the positive imaginary  $\gamma$  axis. As one approaches the positive imaginary  $\gamma$  axis from the right, one crosses all Stokes curves of positive  $y$ . As one approaches it from the left, one crosses all Stokes curves

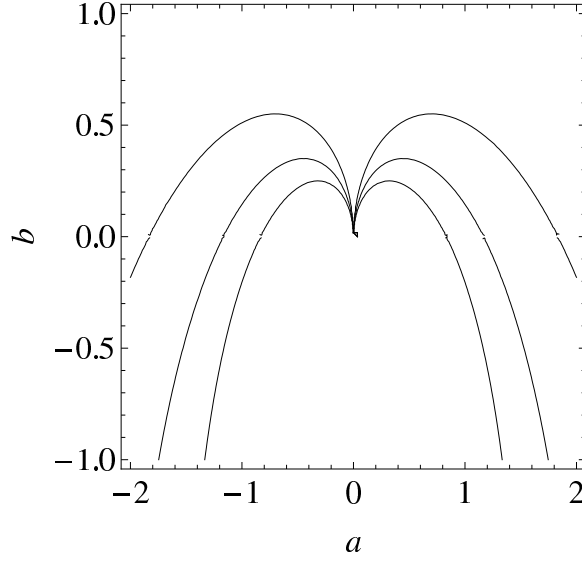


Figure 14: The Stokes curves across which jumping occurs for the trefoil knot divide the complex  $\gamma$  plane as shown here. Sketched here are the curves corresponding to  $y = \pm 5/6, \pm 7/6$ , and  $\pm 11/6$ . (All the other curves for larger or smaller values of  $|y|$  are obtained from these by suitable scalings of the  $\gamma$ -plane.) Drawn are only the parts of each curve that actually produce jumping; these portions meet only at  $\gamma = 0$ , where they all join.

of negative  $y$ . A first consequence is that, precisely on the positive imaginary axis, the representation of the Chern-Simons integration cycle as a sum over Lefschetz thimbles is an infinite sum. A second consequence is that, unless we can consider the Lefschetz thimble  $\mathcal{J}_y$  to be equivalent to  $\mathcal{J}_{-y}$ , this representation of the integration cycle jumps in crossing the positive imaginary  $\gamma$  axis. Indeed, in a complete circuit around the point  $\gamma = 0$ , one crosses exactly once each of the relevant Lefschetz thimbles of fig. 14. This will produce a monodromy of infinite order, and hence an essential singularity in the integral, unless  $\mathcal{J}_y$  is equivalent to  $\mathcal{J}_{-y}$ .

Can we consider  $\mathcal{J}_y$  and  $\mathcal{J}_{-y}$  to be equivalent? Referring back to (5.29), we see that  $y \rightarrow -y$  amounts to  $s \rightarrow s$  with an integer change in  $r$ . In other words, it amounts to a change in the choice of branch of  $\ln \mathbf{m}$ . The branch of  $\ln \mathbf{m}$  is irrelevant as long as  $n \in \mathbb{Z}$  and we choose not to replace the loop space of  $G/T_{\mathbb{C}}$  by its universal cover. This appears to be the right sense in which the integration cycle is univalent near  $\gamma = 0$ , but clearly, a study of more examples is needed. For  $n \notin \mathbb{Z}$ , the monodromy of the integration cycle is unavoidable and an essential singularity of the integral at  $\gamma = 0$  should be expected.

### 5.3 The Figure-Eight Knot

Here we will consider in a similar way the figure-eight knot (also known as the  $4_1$  knot), which we denote  $K$ . The fundamental group of  $S^3 \setminus K$  again has two generators  $\mathbf{x}, \mathbf{y}$ , either of them conjugate to the meridian  $\widehat{\mathbf{m}}$ , with one relation. The relation is  $\omega \mathbf{x} = \mathbf{y} \omega$ , with  $\omega = \mathbf{x} \mathbf{y}^{-1} \mathbf{x}^{-1} \mathbf{y}$ .

For a generic eigenvalue  $\mathbf{m}$  of the meridian, there are up to gauge transformation three  $SL(2, \mathbb{C})$  flat connections on  $S^3 \setminus K$ . As usual, one is an abelian flat connection  $\mathcal{A}_{\text{ab}}$ , whose monodromies are up to conjugacy  $\mathbf{x} = \mathbf{y} = \text{diag}(\mathbf{m}, \mathbf{m}^{-1})$ . In addition, there are two non-abelian flat connections with monodromies

$$\mathbf{x} = \begin{pmatrix} \mathbf{m} & 1 \\ 0 & \mathbf{m}^{-1} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{m} & 0 \\ -d & \mathbf{m}^{-1} \end{pmatrix}, \quad (5.40)$$

with

$$d = \frac{1}{2} \left( \mathbf{m}^2 + \mathbf{m}^{-2} - 3 \pm \sqrt{(\mathbf{m}^2 + \mathbf{m}^{-2} + 1)(\mathbf{m}^2 + \mathbf{m}^{-2} - 3)} \right). \quad (5.41)$$

The longitude is  $\widehat{\mathbf{l}} = \mathbf{x} \mathbf{y}^{-1} \mathbf{x} \mathbf{y} \mathbf{x}^{-2} \mathbf{y} \mathbf{x} \mathbf{y}^{-1} \mathbf{x}^{-1}$ , leading to

$$\mathbf{l} = \frac{1}{2} (\mathbf{m}^4 - \mathbf{m}^2 - 2 - \mathbf{m}^{-2} + \mathbf{m}^{-4}) \pm \frac{\mathbf{m}^2 - \mathbf{m}^{-2}}{2} \sqrt{(\mathbf{m}^2 + \mathbf{m}^{-2} + 1)(\mathbf{m}^2 + \mathbf{m}^{-2} - 3)}. \quad (5.42)$$

We denote as  $\mathcal{A}_{\pm}$  a flat connection with the monodromies just indicated. These are the  $SL(2, \mathbb{C})$  flat connections that, if  $\mathbf{m} = \pm 1$ , are related to the complete hyperbolic metric on  $S^3 \setminus K$ .

The longitude  $\mathbf{l}$  is a zero of the  $\mathbf{A}$  polynomial, which is

$$\mathbf{A}(\mathbf{l}, \mathbf{m}) = \mathbf{l} - (\mathbf{m}^4 - \mathbf{m}^2 - 2 - \mathbf{m}^{-2} + \mathbf{m}^{-4}) + \mathbf{l}^{-1}. \quad (5.43)$$

Apart from the standard symmetries  $\mathbf{m} \rightarrow -\mathbf{m}$  and  $\mathbf{m} \rightarrow \mathbf{m}^{-1}$ , this has an additional symmetry  $\mathbf{l} \rightarrow \mathbf{l}^{-1}$ , which exchanges the two branches of the square root in (5.42). This symmetry reflects the existence of a diffeomorphism  $\tau$  of  $S^3$  that maps  $K$  to itself, reversing the orientation of  $K$  but preserving the orientation of its normal bundle. ( $\tau$  acts on the above presentation of the fundamental group by  $\mathbf{x} \leftrightarrow \mathbf{y}$ .) Such a symmetry maps  $\mathbf{m} \rightarrow \mathbf{m}$ ,  $\mathbf{l} \rightarrow \mathbf{l}^{-1}$ . It reverses the orientation of  $S^3$ , so it changes the sign of the Chern-Simons function  $W(\mathcal{A})$  and is not a symmetry of Chern-Simons gauge theory. Nevertheless, it will help us in understanding the behavior of the theory. A knot for which such a symmetry exists is called amphichiral. Actually, the figure-eight knot is amphichiral in the strong sense that  $S^3$  admits another diffeomorphism  $\widetilde{\tau}$  that maps  $K$  to itself (pointwise) while reversing the orientation of its normal bundle. (It acts on the presentation of the fundamental group by  $\mathbf{x} \rightarrow \mathbf{y}^{-1}$ ,  $\mathbf{y} \rightarrow \mathbf{x}^{-1}$ .) Again,  $\widetilde{\tau}$  reverses the orientation of  $S^3$  and reverses the sign of the Chern-Simons invariant. It acts by  $\mathbf{m} \rightarrow \mathbf{m}^{-1}$ ,  $\mathbf{l} \rightarrow \mathbf{l}$ . (We leave it to the reader to verify that the unknot is

a simpler example of a knot in  $S^3$  that admits diffeomorphisms of both of these types.) The product  $\tau\tilde{\tau}$  preserves the Chern-Simons invariant and acts by  $\mathbf{m} \rightarrow \mathbf{m}^{-1}$ ,  $\mathbf{l} \rightarrow \mathbf{l}^{-1}$ .

Unlike the trefoil, where all of the interesting action occurs for  $|\mathbf{m}| = 1$ , for the figure-eight knot both the real  $\mathbf{m}$  axis and the circle  $|\mathbf{m}| = 1$  are worthy of careful study. The eigenvalues of an element of  $SL(2, \mathbb{R})$  are either real or of modulus 1, so in general flat connections with  $SL(2, \mathbb{R})$  structure group may have  $|\mathbf{m}| = 1$  (as for the trefoil) or real  $\mathbf{m}$  (as happens for the figure-eight knot). The eigenvalues of an element of  $SU(2)$  are of modulus 1, so flat connections with  $SU(2)$  structure group will arise only for  $|\mathbf{m}| = 1$ .

### 5.3.1 Real $\mathbf{m}$

Suppose first that  $\mathbf{m}$  is real and that

$$\mathbf{m}^2 + \mathbf{m}^{-2} > 3. \quad (5.44)$$

Then by inspection, for both flat bundles  $\mathcal{A}_{\pm}$ , the matrices  $\mathbf{x}$  and  $\mathbf{y}$  are real, so the entire representation of the fundamental group has its image in  $SL(2, \mathbb{R})$ . The diffeomorphism  $\tau$  exchanges  $\mathcal{A}_{+}$  with  $\mathcal{A}_{-}$  (since it maps  $\mathbf{l} \rightarrow \mathbf{l}^{-1}$ ). It leaves  $\mathcal{A}_{\text{ab}}$  invariant.

The region with  $\mathbf{m}$  real and

$$\mathbf{m}^2 + \mathbf{m}^{-2} < 3 \quad (5.45)$$

is also easy to understand qualitatively. The flat connections  $\mathcal{A}_{+}$  and  $\mathcal{A}_{-}$  are still exchanged by  $\tau$ . Now, however, they are no longer real (since the square roots in the monodromy formulas are imaginary) and are exchanged by complex conjugation. In fact, the region  $\mathbf{m}^2 + \mathbf{m}^{-2} < 3$  includes the points  $\mathbf{m} = \pm 1$  where  $\mathcal{A}_{+}$  and  $\mathcal{A}_{-}$  are the  $SL(2, \mathbb{C})$  flat connections that are related to a hyperbolic metric on  $S^3 \setminus K$ .

As one might expect, among the most interesting points on the real  $\mathbf{m}$  axis are the points with

$$\mathbf{m}^2 + \mathbf{m}^{-2} = 3. \quad (5.46)$$

For such  $\mathbf{m}$ , since  $d$  vanishes, the matrices  $\mathbf{x}$  and  $\mathbf{y}$  are both upper triangular. The flat bundles  $\mathcal{A}_{+}$  and  $\mathcal{A}_{-}$  become equivalent to each other. They are unstable upper triangular deformations of the abelian flat bundle  $\mathcal{A}_{\text{ab}}$ .  $\mathcal{A}_{\text{ab}}$ , just as in (5.18), is a diagonal flat connection

$$\begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}. \quad (5.47)$$

It has automorphism group  $\mathbb{C}^*$ ; this symmetry group reduces to  $\mathbb{R}^* \cong SO(1, 1)$  if we consider  $\mathcal{A}_{\text{ab}}$  (for real  $\mathbf{m}$ ) as a flat bundle with structure group  $SL(2, \mathbb{R})$ . The symmetry extends to  $O(1, 1)$  if we include the outer automorphism of  $SL(2, \mathbb{R})$  corresponding to conjugation by

$\text{diag}(1, -1)$  or  $\text{diag}(i, -i)$ , and we will do so.<sup>31</sup>  $\mathcal{A}_{\text{ab}}$  has no deformations as a flat bundle with specified  $\mathfrak{m}$ , but as in (5.19), it has a one-dimensional space of deformations of upper triangular form

$$\begin{pmatrix} 0 & u \\ 0 & 0 \end{pmatrix}. \quad (5.48)$$

Here  $u$  is a real parameter if we consider deformations in  $SL(2, \mathbb{R})$ . The value of  $u$ , assuming it is not zero, does not matter modulo the action of  $O(1, 1)$ , and the upper triangular deformation gives us the triangular flat bundle  $\mathcal{A}_+$  or  $\mathcal{A}_-$ . Just as in (5.20),  $\mathcal{A}_{\text{ab}}$  has a similar lower triangular deformation, schematically

$$\begin{pmatrix} 0 & 0 \\ -v & 0 \end{pmatrix}. \quad (5.49)$$

If we take  $v \neq 0$  but  $u = 0$ , we get a lower triangular flat bundle, and again the value of  $v$  is inessential assuming that it is nonzero.

As in our discussion of the trefoil, there is no deformation with  $u, v$  nonzero and  $\mathfrak{m}^2 + \mathfrak{m}^{-2} = 3$ , since a flat  $SL(2, \mathbb{C})$  bundle at such a value of  $\mathfrak{m}$  is triangular (or abelian). If we do turn on both  $u$  and  $v$ , the  $O(1, 1)$ -invariant parameter is the product  $uv$ . One might expect as in the discussion of the trefoil that we would have  $uv \sim \epsilon$  where  $\epsilon = \mathfrak{m}^2 + \mathfrak{m}^{-2} - 3$ . However, in (5.40) the upper triangular parameter is 1 and the lower triangular parameter is  $-d$ , so this case corresponds to  $uv \sim -d$ . Moreover, from (5.41), we have  $d \sim \pm\sqrt{\epsilon}$  near  $\epsilon = 0$ . Critical points therefore are at  $uv = 0$  or  $\pm\sqrt{\epsilon}$ . Apart from a multiplicative constant, the function with this critical behavior is

$$f(u, v) = (uv)^3/3 - \epsilon uv. \quad (5.50)$$

The oscillatory integral associated to this function was studied in section 3.4.3.

Why have we arrived at this peculiar structure, with no  $(uv)^2$  term in  $f(u, v)$ ? The answer has to do with amphichirality. As  $\tau$  leaves fixed  $\mathcal{A}_{\text{ab}}$  and exchanges  $\mathcal{A}_{\pm}$ , at a point at which one of these becomes triangular and meets  $\mathcal{A}_{\text{ab}}$ , the other must do so as well. This structure is lost if we add a  $(uv)^2$  term to  $f$ . More fundamentally,  $\tau$  reverses the sign of the Chern-Simons function and hence of  $f$ , which describes its critical behavior. The fact that  $\tau$  exchanges  $\mathcal{A}_{\pm}$  implies that if we lift the action of  $\tau$  to the flat bundle with connection  $\mathcal{A}_{\text{ab}}$  in such a way that  $\tau(u) = u$ , then  $\tau(v) = -v$ . (By composing with the outer automorphism of  $SL(2, \mathbb{R})$ , one can also choose a lifting so that  $\tau(u) = -u$ ,  $\tau(v) = v$ .) Thus,  $\tau$  reverses the sign of  $uv$ . Since  $\tau$  reverses the sign of both  $uv$  and  $f$ ,  $f$  is an odd function of  $uv$ , and this accounts for the structure that we have found.

---

<sup>31</sup>In  $SL(2, \mathbb{R})$  gauge theory, one would consider two critical points that differ by this outer automorphism to be inequivalent. But in  $SL(2, \mathbb{C})$  gauge theory, or analytic continuation of  $SU(2)$  gauge theory, which are our focus in this paper, they are equivalent, since  $\text{diag}(i, -i)$  is an element of  $SU(2)$ .

As in our analysis of the trefoil, some of the Stokes phenomena of Chern-Simons theory of the figure-eight knot can be described in terms of critical behavior of the function  $f$ . However, for brevity, the only issue that we will discuss here will be the large  $n$  behavior of the colored Jones polynomial near  $\epsilon = 0$ , that is, near  $\mathbf{m}^2 + \mathbf{m}^{-2} = 3$ . First we simply set  $\epsilon = 0$ . The parameter  $\lambda$  in eqn. (3.65) corresponds in Chern-Simons theory to  $n$  (or equivalently  $k$ , since the ratio  $n/k$  is fixed when we specify  $\epsilon$ ). So according to (3.65),

$$\frac{J_n(k)|_{\epsilon=0}}{J_n(k)|_{\epsilon<0}} \sim n^{2/3}, \quad n \rightarrow \infty. \quad (5.51)$$

Accordingly, if we normalize the colored Jones polynomial of the figure-eight knot so that it is of order 1 for  $\epsilon < 0$ , then it is of order  $n^{2/3}$  for  $\epsilon = 0$ . This is a known result ([16], Theorem 1.1). Instead of simply setting  $\epsilon = 0$ , we can consider the limit  $n \rightarrow \infty$ ,  $\epsilon \rightarrow 0$  with  $n^{2/3}\epsilon$  fixed. The singular contribution to the colored Jones polynomial in this limit follows from (3.64) and is an Airy function (times  $n^{2/3}$  in the usual mathematical normalization), as one can see by introducing polar coordinates in the  $xy$  plane.

### 5.3.2 $\mathbf{m}$ Of Modulus 1

Now let us analyze what happens for  $|\mathbf{m}| = 1$ . When are the flat connections  $\mathcal{A}_{\pm}$  unitary? For this to be the case,  $\mathbf{l}$  must be of modulus 1. From (5.42), the condition for this (assuming that  $|\mathbf{m}| = 1$ ) is

$$\mathbf{m}^2 + \mathbf{m}^{-2} + 1 \leq 0. \quad (5.52)$$

Conversely, if this condition is obeyed and  $|\mathbf{m}| = 1$ , one can explicitly conjugate the flat bundle with monodromies (5.40) into  $SU(2)$  by the same steps that we used in discussing the trefoil knot.

For  $|\mathbf{m}| = 1$  and  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 > 0$ , the bundles  $\mathcal{A}_{\pm}$  have structure group  $SL(2, \mathbb{C})$ . A reduction to  $SL(2, \mathbb{R})$  cannot occur because  $\widehat{\mathbf{m}}$  has eigenvalues of unit modulus while  $\widehat{\mathbf{l}}$  has real eigenvalues; commuting (and noncentral) elements of  $SL(2, \mathbb{R})$  do not have this property. Also, the region  $|\mathbf{m}| = 1$ ,  $\mathbf{m}^2 + \mathbf{m}^{-2} > 1$  again contains the points  $\mathbf{m} = \pm 1$  where the connections  $\mathcal{A}_{\pm}$  are  $SL(2, \mathbb{C})$  connections related to the hyperbolic metric on  $S^3 \setminus K$ .

As usual, we have to look more closely at what happens for  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 = 0$ . It is convenient to set  $\epsilon = -\mathbf{m}^2 - \mathbf{m}^{-2} - 1$ , so that  $\epsilon$  is real for  $|\mathbf{m}| = 1$  and the flat connections  $\mathcal{A}_{\pm}$  have  $SU(2)$  structure group precisely when  $\epsilon \geq 0$ . Unlike our previous examples, the nonabelian flat connections  $\mathcal{A}_{\pm}$  remain irreducible at those points and as a result the abelian flat connection  $\mathcal{A}_{\text{ab}}$  will play no role in critical behavior near  $\epsilon = 0$ . All that happens when  $\epsilon = 0$  is that the two critical points  $\mathcal{A}_{\text{ab}}$  become coincident. Moreover, near  $\epsilon = 0$ , the difference between the two critical points is of order  $\epsilon^{1/2}$ ; this is evident from the formulas



(5.40) for the monodromies. Finally, the critical points are real (in the natural real structure of  $SU(2)$  gauge theory) if  $\epsilon > 0$ . These are the properties of the familiar function

$$f(x) = x^3/3 - \epsilon x \quad (5.53)$$

associated to the Airy integral. This function describes the critical behavior of the Chern-Simons function  $W(\mathcal{A})$  near the points  $\epsilon = 0$ , and hence in studying Stokes phenomena, we will be able to borrow results from section 3.4.1.

For now, we simply describe qualitatively the behavior of the colored Jones polynomial  $J_n(q)$  for real  $\epsilon$  near  $\epsilon = 0$ . First we consider the limit  $n \rightarrow \infty$  with  $\epsilon$  fixed:

(1) For  $\epsilon < 0$ , the only real critical point is the abelian one. Its contribution to the path integral, for any knot, is of order  $1/\sqrt{n}$ , so we have

$$J_n(q) \sim \frac{C}{n^{1/2}}, \quad \epsilon < 0 \quad (5.54)$$

with a constant  $C$ . (Actually, in a more precise description,  $C$  is the first term in an asymptotic expansion  $C + C'/n + \dots$  in integer powers of  $1/n$ .)

(2) For  $\epsilon > 0$ , the flat bundles  $\mathcal{A}_\pm$  appear as real critical points in the path integral. They give oscillatory contributions, but without the factor  $1/\sqrt{n}$  that appears in the contribution of  $\mathcal{A}_{\text{ab}}$ . Including also the abelian contribution, the integral behaves as

$$J_n(\epsilon) \sim \frac{C}{n^{1/2}} + C_+ \exp(inf_+) + C_- \exp(inf_-), \quad \epsilon > 0 \quad (5.55)$$

with constants  $C_\pm$  (which again are the leading terms of asymptotic series). Here we have written the action  $\mathcal{I}(\mathcal{A}_\pm)$  as  $inf_\pm(\epsilon)$ , where  $f_\pm(\epsilon)$  is independent of  $n$ . We have  $f_\pm(0) = 0$ , since  $\mathcal{A}_\pm$  coincide with  $\mathcal{A}_{\text{ab}}$  at  $\epsilon = 0$ . Also, as explained in section 5.3.3,  $f_+(\epsilon) = -f_-(\epsilon)$  for real  $\epsilon$ .

(3) What happens at  $\epsilon = 0$ ? Typically the  $n$ -dependence of the integral over any one mode of the path integral is that of a Gaussian integral  $\int dx \exp(ix^2) \sim 1/n^{1/2}$ . For the path integral around an irreducible flat connection with vanishing cohomology (such as  $\mathcal{A}_\pm$  at  $\epsilon \neq 0$ ), these factors, after taking into account the regularization and the volume of the gauge group, ultimately cancel in a complete evaluation of the path integral. This is why the oscillatory terms in (5.55) do not have a power of  $n$  as a prefactor. But at  $\epsilon = 0$ , the integral over  $x$ , rather than being Gaussian, becomes  $\int dx \exp(ix^3) \sim 1/n^{1/3} = n^{1/6}/n^{1/2}$ . The denominator participates in the usual cancellation, but the numerator does not, so we expect

$$J_n(0) \sim n^{1/6}. \quad (5.56)$$

(In the usual mathematical normalization, one divides by the path integral (5.3) of the unknot, and then the colored Jones polynomial grows as  $n^{2/3}$ .)

So far we have described what happens for  $n \rightarrow \infty$  with fixed  $\epsilon$ . If instead we consider the limit  $n \rightarrow \infty$  with  $n^{2/3}\epsilon$  fixed, then in view of (3.55), the singular behavior of the colored Jones polynomial is an Airy function (times  $n^{2/3}$  in the usual mathematical normalization).

### 5.3.3 The Real Part Of The Chern-Simons Function

For the figure-eight knot, we will study only the simplest properties of the Chern-Simons function.

First we focus on the region  $|\mathbf{m}| = 1$ . We denote as  $\mathcal{C}$  the operation of complex conjugation of a flat  $SL(2, \mathbb{C})$  bundle.  $\mathcal{C}$  maps  $(\mathbf{m}, \mathbf{l})$  to  $(\bar{\mathbf{m}}, \bar{\mathbf{l}})$ . Let us compose this operation with the symmetry  $\tilde{\tau}$ , which maps  $(\mathbf{m}, \mathbf{l})$  to  $(\mathbf{m}^{-1}, \mathbf{l})$ . The composite acts by  $(\mathbf{m}, \mathbf{l}) \rightarrow (\bar{\mathbf{m}}^{-1}, \bar{\mathbf{l}})$ . If  $|\mathbf{m}| = 1$ , then  $\bar{\mathbf{m}}^{-1} = \mathbf{m}$  and the action is  $(\mathbf{m}, \mathbf{l}) \rightarrow (\mathbf{m}, \bar{\mathbf{l}})$ .

There are two somewhat different cases:

(1) If  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 < 0$ , the two flat bundles are  $SU(2)$  bundles with complex conjugate values of  $\mathbf{l}$ . The composite  $\mathcal{C}\tilde{\tau}$  exchanges the two values of  $\mathbf{l}$ , so it exchanges the two flat bundles.

(2) If  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 > 0$ , the two flat bundles are  $SL(2, \mathbb{C})$  bundles with real and unequal values of  $\mathbf{l}$ . The composite  $\mathcal{C}\tilde{\tau}$  leaves fixed the two values of  $\mathbf{l}$ , so it also leaves fixed the isomorphism class of each of these two flat bundles.

Now let us analyze the Chern-Simons invariant  $W(\mathcal{A})$  in the two cases.  $\mathcal{C}$  transforms  $W$  to its complex conjugate, while  $\tilde{\tau}$  (because it reverses the orientation of  $S^3$ ) maps it to  $-W$ . So the composite maps  $W$  to  $-\bar{W}$ .

In case (1),  $W$  is real, since in general the Chern-Simons invariant of an  $SU(2)$  connection is real. So  $W$  transforms simply to  $-W$  under the exchange of the two flat bundles. Thus they have equal and opposite real values of  $W$ .

In case (2),  $W$  is complex. The fact that the two flat bundles are invariant under  $\mathcal{C}\tilde{\tau}$  tells us that their Chern-Simons functions are invariant under  $W \rightarrow -\bar{W}$ . This tells us nothing about  $\text{Im } W$ , but it implies that  $\text{Re } W = -\text{Re } W$ . Before concluding that  $\text{Re } W = 0$ , we should recall that  $\text{Re } W$  is only well-defined modulo  $2\pi$ . So the correct conclusion is only that  $\text{Re } W$  vanishes mod  $\pi$ .

In case (2), we are left with two possible cases:  $\text{Re } W$  may vanish mod  $2\pi$  or may equal  $\pi$  mod  $2\pi$ . The region  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 > 0$ ,  $|\mathbf{m}| = 1$ , has two connected components, as  $\mathbf{m}$  may

be near either 1 or  $-1$ . It turns out that  $\operatorname{Re} W = 0 \bmod 2\pi$  if  $\mathbf{m}$  is near 1, and  $\operatorname{Re} W = \pi \bmod 2\pi$  if  $\mathbf{m}$  is near  $-1$ .

To show this, we observe first that if  $\mathbf{m}^2 + \mathbf{m}^{-2} + 3 = 0$ , then the flat bundle is triangular, which guarantees that  $W = 0 \bmod 2\pi$ . Let  $\mathbf{m}_0$  be a positive root of the equation  $\mathbf{m}^2 + \mathbf{m}^{-2} + 3 = 0$ . Then  $-\mathbf{m}_0$  is a negative root. We can calculate  $W(\mathcal{A})$  for  $\mathbf{m} = 1$  by integrating eqn. (4.42) along the real  $\mathbf{m}$  axis from  $\mathbf{m}_0$  to 1. We get

$$W(\mathcal{A}; \mathbf{m} = 1) = W(\mathcal{A}; \mathbf{m} = \mathbf{m}_0) + (1/\pi) \int_{\mathbf{m}_0}^1 \ln \mathbf{m} \, d \ln \mathfrak{l}. \quad (5.57)$$

Along the integration path,  $\ln \mathbf{m}$  is real. On the other hand, according to eqn. (5.42),  $\mathfrak{l}$  is of modulus 1 for  $\mathbf{m}$  real,<sup>32</sup> so  $\ln \mathfrak{l}$  is imaginary. Hence, the integral is imaginary. Since the integration constant  $W(\mathcal{A}; \mathbf{m} = \mathbf{m}_0)$  vanishes, we learn that

$$\operatorname{Re} W(\mathcal{A}; \mathbf{m} = 1) = 0 \bmod 2\pi. \quad (5.58)$$

Similarly we have  $W(\mathcal{A}; \mathbf{m} = -1) = W(\mathcal{A}; \mathbf{m} = -\mathbf{m}_0) + (1/\pi) \int_{-\mathbf{m}_0}^{-1} \ln \mathbf{m} \, d \ln \mathfrak{l}$ , and again the integration constant vanishes. But now we get a different result because the imaginary part of  $\ln \mathbf{m}$  along the negative  $\mathbf{m}$  axis is equal to  $\pi$ . From (5.42), we see that as  $\mathbf{m}$  varies from  $-\mathbf{m}_0$  to  $-1$ ,  $\mathfrak{l}$  makes a half-circuit around the unit circle from 1 to  $-1$ . So  $\ln \mathfrak{l}$  changes by  $i\pi$  and finally we get

$$\operatorname{Re} W(\mathcal{A}; \mathbf{m} = -1) = \pi \bmod 2\pi. \quad (5.59)$$

Now let us discuss the behavior for  $\mathbf{m} = \exp(i\pi\gamma)$  as a function of real  $\gamma$  with  $|\gamma| \leq 1$ . The condition  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 < 0$  amounts to  $1/3 < |\gamma| < 2/3$ . So case (2) in the above analysis corresponds to  $|\gamma| < 1/3$  or  $|\gamma| > 2/3$ . In either of those regions,  $\operatorname{Re} W$  can only take the values 0 or  $\pi$ , so it is constant by continuity.<sup>33</sup> Hence  $\operatorname{Re} W = 0$  if  $|\gamma| \leq 1/3$ , and  $\operatorname{Re} W = \pi$  if  $|\gamma| \geq 2/3$ .

For  $1/3 \leq |\gamma| \leq 2/3$ , there are up to isomorphism two flat  $SU(2)$  connections, which we denote as  $A_{\pm}$ . At  $|\gamma| = 1/3$ ,  $A_+$  and  $A_-$  both coincide with the flat  $SL(2, \mathbb{C})$  connection considered in the last paragraph. So in that case,  $\operatorname{Re} W = 0 \bmod 2\pi$ . It is convenient to lift the Chern-Simons invariants so that  $A_{\pm}$  have  $\operatorname{Re} W = 0$  at  $|\gamma| = 1/3$ .

Now let us follow the two  $SU(2)$  flat connections  $A_+$  and  $A_-$  continuously from  $|\gamma| = 1/3$  to  $|\gamma| = 2/3$ . At  $|\gamma| = 2/3$ , we can use (5.59), along with the constancy of  $\operatorname{Re} W$  for  $|\gamma| \geq 2/3$ , to deduce that  $W(A_+)$  and  $W(A_-)$  are each congruent to  $\pi \bmod 2\pi$ . Since we also showed above that  $W(A_+) = -W(A_-)$ , it follows that for some integer  $w$ , we have  $W(A_{\pm}) = \pm(2w + 1)\pi$  at  $|\gamma| = 2/3$ .

---

<sup>32</sup>Fundamentally, this reflects the fact that for  $\mathbf{m}$  real, the two flat bundles are invariant under the composite  $C\tau$ , which leaves real  $\mathbf{m}$  fixed while mapping  $\mathfrak{l}$  to  $\bar{\mathfrak{l}}^{-1}$ .

<sup>33</sup>This constancy is also a direct consequence of eqn. (4.42). In the region in question,  $\ln \mathbf{m}$  is imaginary and  $\ln \mathfrak{l}$  is real, so the real part of  $\ln \mathbf{m} \, d \ln \mathfrak{l}$  vanishes.

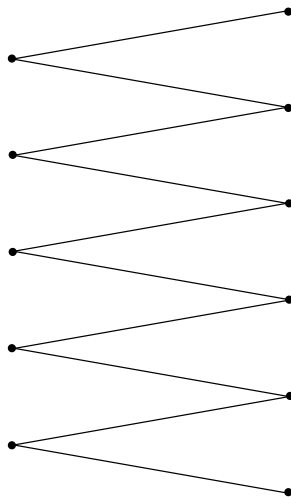


Figure 15: Qualitative behavior of the conserved quantity  $\text{Im } I/k$  for real critical points in the region  $1/3 \leq |\gamma| \leq 2/3$ . The structure is in perfect parallel with fig. 9 of section 3.5.2, except now  $\gamma$  is plotted on the horizontal axis. At  $|\gamma| = 1/3$  (left), there are a pair of real critical points for every value of  $\text{Im } \mathcal{I}/k$  of the form  $2\pi w$ ,  $w \in \mathbb{Z}$ . Increasing  $|\gamma|$ ,  $\text{Im } \mathcal{I}/k$  increases for one of these critical points and reduces for the other. At  $|\gamma| = 2/3$  (right), the critical points recombine in pairs, but now with  $\text{Im } \mathcal{I}/k$  of the form  $(2w + 1)\pi$ ,  $w \in \mathbb{Z}$ .

It can be shown by integrating (4.42) that  $w = 1$ . (This is greatly facilitated by symmetrizing the integrand between  $\gamma$  and  $1 - \gamma$  before attempting to integrate.) Now let us discuss what happens when we include all possible lifts of  $W(A_{\pm})$ .

For  $|\gamma| < 1/3$ , critical points come in complex conjugate pairs. In passing through  $|\gamma| = 1/3$ , a pair of complex conjugate critical points meet and recombine into two inequivalent real critical points  $A_{\pm}$ . Actually, this occurs infinitely many times, once corresponding to each complex lift of  $W$ .

Increasing  $\gamma$  past  $|\gamma| = 2/3$ , the reverse process occurs: a pair of inequivalent real critical points recombine into a complex conjugate pair of critical points. Again this actually occurs infinitely many times.

The flow in  $W$  between  $|\gamma| = 1/3$  and  $|\gamma| = 2/3$  leads to a crucial detail. Actually, this detail also arose in analytic continuation of the Bessel function in section 3.5, and was illustrated in fig. 9. That figure is repeated for convenience in fig. 15, with a caption adapted to the present problem. The point is that a pair of real critical points that are “born” at  $|\gamma| = 1/3$  with the same value of  $W$  no longer have the same value of  $W$  by the time they “arrive” at  $|\gamma| = 2/3$ . Rather they have values of  $W$  that differ by  $2\pi$ . The pairs of critical points that separate at  $|\gamma| = 1/3$  are not the same as the pairs that rejoin at  $|\gamma| = 2/3$ .

Precisely the same phenomenon led in section 3.4.2 to exponential growth of the analytically continued Bessel function for negative and non-integer values of  $\kappa$ . It should therefore come as no surprise that this mechanism will lead to exponential growth of the function  $J_n(q)$  associated to the figure-eight knot in the region  $|\gamma| > 2/3$ , as long as the Chern-Simons level  $k$  is not an integer.

### 5.3.4 The Imaginary Part Of $W$

We will now discuss much more briefly the simplest properties of  $\text{Im } W$  along the unit circle  $|\mathbf{m}| = 1$ .

For  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 \leq 0$ , the nonabelian flat connections are  $SU(2)$ -valued flat connections  $A_{\pm}$ . As we have already noted, like any  $SU(2)$ -valued connections, these have  $\text{Im } W = 0$ .

For  $\mathbf{m}^2 + \mathbf{m}^{-2} + 1 > 1$ , there are a pair of  $SL(2, \mathbb{C})$  flat connections  $\mathcal{A}_{\pm}$ , exchanged by complex conjugation. Since  $\text{Im } W$  is odd under complex conjugation, we have  $\text{Im } W(\mathcal{A}_+) = -\text{Im } W(\mathcal{A}_-)$ .

The most important property of the invariant  $\text{Im } W$  is its relation to the hyperbolic volume of the knot complement  $S^3 \setminus K$ .  $S^3 \setminus K$  admits a complete hyperbolic metric, unique up to diffeomorphism. Though this metric is complete, it has the property that the volume is convergent.<sup>34</sup> We denote this volume as  $V$ .

Like a hyperbolic metric on any three-manifold, the complete hyperbolic metric on  $S^3 \setminus K$  determines a flat  $PSL(2, \mathbb{C})$  bundle on this manifold. Upon picking a spin structure on  $S^3 \setminus K$ , the flat  $PSL(2, \mathbb{C})$  bundle can be lifted to a flat  $SL(2, \mathbb{C})$  bundle. The conjugacy class of the meridian for this flat bundle is

$$\widehat{\mathbf{m}} = \pm \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad (5.60)$$

where the sign depends on the choice of spin structure (in other words, the two flat bundles with different signs of  $\mathbf{m}$  differ by twisting by a flat line bundle of order 2 – the operation considered in section 4.2.7). The flow equations in the presence of such unipotent monodromy were discussed in section 4.2.2; for our present purposes, all we need to know is that the

---

<sup>34</sup>Near  $K$ , the metric can be modeled by  $ds^2 = z^2(dx^2 + dy^2) + dz^2/z^2$ , where  $x$  and  $y$  parametrize a two-torus and  $K$  is at  $z = 0$ . Thus the region near  $K$  is the “infrared” region in the context of the AdS/CFT correspondence. This description can be used to compute the conjugacy class of  $\widehat{\mathbf{m}}$  given in eqn. (5.60) below. Both  $\widehat{\mathbf{m}}$  and  $\widehat{\mathbf{l}}$  arise in this description from shifts in  $x$  and  $y$ ; such shifts correspond to unipotent elements of  $PSL(2, \mathbb{C})$ .

eigenvalue of  $\widehat{\mathbf{m}}$  is  $\mathbf{m} = \pm 1$ . On the other hand, from eqn. (5.42), it follows that<sup>35</sup>  $\mathfrak{l} = -1$ , for either sign of  $\mathbf{m}$ .

The relation between the hyperbolic volume and the Chern-Simons invariant is simply that, if we label  $\mathcal{A}_{\pm}$  correctly,

$$\operatorname{Re}(iW(\mathcal{A}_{\pm}; \mathbf{m} = \pm 1)) = \pm \frac{V}{2\pi}. \quad (5.61)$$

This means that  $\operatorname{Im} W(\mathcal{A}_{\pm}; \mathbf{m} = \pm 1) = \mp V/2\pi$ . The sign may look unnatural, but it ensures that  $\mathcal{A}_{\pm}$  make exponentially growing and decaying contributions to the path integral, respectively; this seems like the most useful mnemonic.

If we depart from  $\mathbf{m} = \pm 1$ , there is no longer a complete hyperbolic metric on  $S^3 \setminus K$  with  $\mathbf{m}$  as the eigenvalue of the meridian. But there is an incomplete hyperbolic metric with this property (it has a conical singularity along  $K$ ), and a statement along the lines of (5.61) remains valid for this incomplete metric. We will not describe details here.

We have explained the qualitative properties of  $\operatorname{Re} W$  and  $\operatorname{Im} W$  that will be of most concern for the limited application that we consider next. We should note, however, that (as summarized in [17], for example) by integrating the usual formula (4.42) and using (5.42), one can express  $W(\mathcal{A})$  for flat connections on the complement of the figure-eight knot in terms of a dilogarithm function.

### 5.3.5 The Volume Conjecture For The Figure-Eight Knot

The volume conjecture for knots emerged from the following observation [2, 3] about the analytically continued invariant  $J_n(q)$  of the figure-eight knot. For  $\gamma$  approaching 1 from a direction away from the real axis, the large  $n$  behavior of  $J_n(q)$  is, roughly speaking,

$$J_n(q) \sim \exp(nV/2\pi), \quad (5.62)$$

where  $V$  is the hyperbolic volume of the knot complement, discussed in section 5.3.4. A more precise statement will emerge below.

A partial explanation of this was given in [6]. It was noted that  $nV/2\pi$  is the value of the exponent  $\mathcal{I} = ikW(\mathcal{A})$  of the analytically-continued Chern-Simons path integral at the complex critical point  $\mathcal{A}_+$ . So an expansion around this critical point would give a

---

<sup>35</sup>From this we can actually give a new demonstration of one result of section 5.3.3, which was that  $\operatorname{Re} W$  shifts by  $\pi$  under  $\mathbf{m} \rightarrow -\mathbf{m}$ . In section 4.2.2, we showed that the shift is  $\Delta W = -i \ln \mathfrak{l}$ , which here equals  $\pi \bmod 2\pi$  as  $\mathfrak{l} = -1$ .

contribution growing as  $\exp(V/2\pi)$ . (In this paper, and subsequent literature, generalizations of the statement to other values of  $\gamma$  were also made. For example, see [3, 5, 9].)

There has always been a puzzle about this partial explanation. For physical values of  $k$  and  $n$  – positive integers – the Chern-Simons path integral is dominated in the semiclassical limit by real critical points. Contributions, if any, of complex critical points are exponentially small. Therefore, at least for positive rational values of  $\gamma = n/k$ , there are no exponentially growing contributions to the path integral. How then can exponential growth be observed just because  $\gamma$  is infinitesimally displaced from 1?

The reader who has paid close attention to sections 3.4.2 and 5.3.4 will hopefully anticipate the answer. Once we analytically continue away from integer values of  $k$ , we need to work not on the space  $\mathcal{U}_{\mathbb{C}}$  of complex-valued connections modulo gauge transformations, but on a cover  $\widehat{\mathcal{U}}_{\mathbb{C}}$  on which the Chern-Simons function  $W(\mathcal{A})$  is  $\mathbb{C}$ -valued (rather than valued in  $\mathbb{C}/2\pi\mathbb{Z}$ ). A given flat connection  $\mathcal{A}$  corresponds to infinitely many critical points that differ only by the choice of a complex lift of  $W(\mathcal{A})$ . When the integration cycle of the analytically continued theory is expressed in terms of Lefschetz thimbles associated to critical points, it is possible for contributions from critical points that differ only in this lift to cancel for integer  $k$ , but give exponential growth as soon as  $k$  is not an integer.

Let us first express in terms of Lefschetz thimbles the integration cycle  $\mathcal{C}$  of Chern-Simons theory for what proves to be the simplest case: positive integer  $k$  and  $n$  with  $1/3 < \gamma < 2/3$ , where as usual  $\gamma = k/n$ . This is easy to do because in this range there are no complex critical points, only real ones. Up to gauge transformation, there is an abelian flat connection  $A_{\text{ab}}$ , and two nonabelian but  $SU(2)$ -valued flat connections  $A_{\pm}$ . As they are all  $SU(2)$ -valued, they all contribute to the integration cycle with a coefficient 1. This cycle is then in an obvious notation

$$\mathcal{C} = \mathcal{J}_{\text{ab}} + \mathcal{J}_{+} + \mathcal{J}_{-}. \quad (5.63)$$

As long as  $k$  and  $n$  are integers, there is also no problem in extending this answer outside the range  $1/3 < \gamma < 2/3$ . At either  $\gamma = 1/3$  or  $\gamma = 2/3$ , there is an Airy-type singularity. The two real flat connections  $A_{\pm}$  recombine to a complex conjugate pair of  $SL(2, \mathbb{C})$ -valued flat connections, which we will call  $\mathcal{A}'_{\pm}$  for  $\gamma < 1/3$  or  $\mathcal{A}''_{\pm}$  for  $\gamma > 2/3$ . We choose the subscript so that, as in (5.61) (where  $\mathbf{m} = \pm 1$  are treated together),  $\text{Re}(iW(\mathcal{A}'_{+})) > 0$  for  $\gamma < 1/3$  and  $\text{Re}(iW(\mathcal{A}''_{+})) > 0$  for  $\gamma > 2/3$ . This ensures that the integrals over the Lefschetz thimbles  $\mathcal{J}'_{-}$  and  $\mathcal{J}''_{-}$  associated to  $\mathcal{A}'_{-}$  and  $\mathcal{A}''_{-}$  are exponentially decaying, while the integrals over the other Lefschetz thimbles  $\mathcal{J}'_{+}$ ,  $\mathcal{J}''_{+}$  associated to  $\mathcal{A}'_{+}$ ,  $\mathcal{A}''_{+}$  are exponentially growing. When we cross the Airy-like singularities at  $\gamma = 1/3$  or  $2/3$ , according to eqn. (3.58), the sum  $\mathcal{J}_{+} + \mathcal{J}_{-}$  recombines to a Lefschetz thimble  $\mathcal{J}'_{-}$  or  $\mathcal{J}''_{-}$ , whose contribution to the path

integral is exponentially decaying. The integration cycle thus becomes

$$\mathcal{C} = \begin{cases} \mathcal{J}_{\text{ab}} + \mathcal{J}'_- & \text{for } \gamma < 1/3 \\ \mathcal{J}_{\text{ab}} + \mathcal{J}''_- & \text{for } \gamma > 2/3. \end{cases} \quad (5.64)$$

Now let us continue away from integer values of  $k$  and  $n$ . Starting in the region  $1/3 < \gamma < 2/3$ , to make an analytic continuation, we must pick lifts of the Chern-Simons invariants  $W(A_{\pm})$ . When we do this, we write, for example,  $\mathcal{J}_{+,w}$  for the Lefschetz thimble associated to  $A_+$  with a Chern-Simons invariant  $2\pi w$ ; we extend this notation to the other flat connections in an obvious way. Here for  $\mathcal{J}_{\pm}$ , the value of  $w \bmod \mathbb{Z}$  depends smoothly on  $\gamma$  (varying from  $0 \bmod \mathbb{Z}$  at  $\gamma = 1/3$  to  $1/2 \bmod \mathbb{Z}$  at  $\gamma = 2/3$ ); for  $\mathcal{J}'_{\pm}$ ,  $w$  takes values in  $\mathbb{Z}$ ; and for  $\mathcal{J}''_{\pm}$ ,  $w$  takes values in  $1/2 + \mathbb{Z}$ .

If we do not want exponential growth for  $n \rightarrow \infty$  and  $\gamma < 1/3$ , we must pick the lifts of  $W(A_+)$  and  $W(A_-)$  to be equal at  $\gamma = 1/3$ ; otherwise, according to (3.58), in continuing past the Airy singularity,  $\mathcal{J}'_+$  as well as  $\mathcal{J}'_-$  will appear in the integration cycle, giving exponential growth. Similarly, if we do not want exponential growth for  $\gamma > 2/3$ , we need  $W(A_+) = W(A_-)$  at  $\gamma = 2/3$ . But it is impossible to obey both of these conditions, since as was illustrated in fig. 15, if  $W(A_+)$  and  $W(A_-)$  are equal at  $\gamma = 1/3$ , then they differ by  $2\pi$  at  $\gamma = 2/3$ , and vice-versa.

Exponential growth is therefore unavoidable, once  $k$  is not an integer, either for  $\gamma < 1/3$  or for  $\gamma > 2/3$ . As explained in section 5.1, we wish to pick an analytic continuation that leads to no exponential growth for  $\gamma < 1/3$ . A further clue is that, to preserve  $\tau$  and  $\tilde{\tau}$  symmetry,  $J_n(q)$  should be real for real  $k$ . (Complex conjugation of the path integral reverses the sign of  $k$ , but as the figure-eight knot is amphichiral, one can compensate for this with the symmetry  $\tau$  or  $\tilde{\tau}$ .) This suggests that we should take the lift of  $\mathcal{A}'_-$  such that  $\text{Re } W(\mathcal{A}'_-) = 0$ . Then the integration cycle for  $\gamma$  just less than  $1/3$  is

$$\mathcal{C} = \mathcal{J}_{\text{ab}} + \mathcal{J}'_{-,0}. \quad (5.65)$$

Continuing past  $\gamma = 1/3$ , cycles  $\mathcal{J}_{+,w}$  and  $\mathcal{J}_{-,w}$  appear, initially with  $w = 0$ . But by the time we reach  $\gamma = 2/3$ , the values are  $w = \pm 1/2$ , as we learned in section 5.3.4. So the integration cycle as  $\gamma$  approaches  $2/3$  from below is

$$\mathcal{C} = \mathcal{J}_{\text{ab}} + \mathcal{J}_{+,1/2} + \mathcal{J}_{-,-1/2}. \quad (5.66)$$

To continue past  $\gamma = 2/3$ , we use eqn. (3.57) to express  $\mathcal{J}_{+,1/2}$  and  $\mathcal{J}_{-,-1/2}$  in terms of corresponding cycles  $\mathcal{J}''_{\pm,w}$ . Since  $w$  does not change in this process, the result is  $\mathcal{J}_{+,1/2} \rightarrow -\mathcal{J}''_{+,1/2} + \mathcal{J}''_{-,1/2}$ , and  $\mathcal{J}_{-,-1/2} \rightarrow \mathcal{J}''_{+,-1/2}$ . The integration cycle for  $\gamma$  just greater than  $1/3$  is therefore

$$\mathcal{C} = \mathcal{J}_{\text{ab}} - \mathcal{J}''_{+,1/2} + \mathcal{J}''_{-,1/2} + \mathcal{J}''_{+,-1/2}. \quad (5.67)$$



Near or on the real  $\gamma$  axis, with  $\text{Re } \gamma > 2/3$ , exponential growth comes from the terms  $\mathcal{J}_{+,-1/2}'' - \mathcal{J}_{+,1/2}''$  in (5.67). Just as in the analytic continuation of the Bessel function, the relative minus sign is the reason that exponential growth is avoided when  $k$  is an integer. For  $\gamma$  near 1, using (5.61), the value of the path integral is

$$J_n(q) \sim \exp(nV/2\pi) (\exp(-i\pi k) - \exp(i\pi k)). \quad (5.68)$$

Here the exponential factors  $\exp(\pm i\pi k)$  reflect the values  $w = \pm 1/2$  for the relevant Lefschetz thimbles. Also, near  $\gamma = 1$ , we have written the exponentially growing factor as  $\exp(nV/2\pi)$  (with  $n$  rather than  $k$  in the exponent) to put the formula in a more recognizable form for some readers.

The expression  $\log J_n(q)/n$  does not have a simple large  $n$  limit if  $\gamma$  is on the real axis, because of the oscillatory factor  $(\exp(-i\pi k) - \exp(i\pi k))$ . As soon as  $\gamma$  is off the real axis, there is a simple limit, but it is discontinuous on the real  $\gamma$  axis. The reason for that is that the large  $n$  limit of  $\log(\exp(-i\pi k) - \exp(i\pi k))/n$  is  $\mp i\pi k/n = \mp i\pi/\gamma$ , depending on the sign of  $\text{Im } k$ .

In the region  $2/3 < \gamma < 4/3$ , the flat connection  $\mathcal{A}_+$ , with its various possible lifts, has the maximum value of  $\text{Re } \mathcal{I}$ , among all flat connections. This ensures that, once the cycles  $\mathcal{J}_{+,\pm 1/2}''$  appear in  $\mathcal{C}$  at  $\gamma = 2/3$ , they cannot disappear at least until  $\gamma = 4/3$ . At  $\gamma = 4/3$ , and more generally at every value of  $\gamma$  of the form  $\pm 1/3 \bmod \mathbb{Z}$ , one meets another family of Airy-type singularities. To analyze the consequences, it may be necessary to take into account Stokes phenomena beyond what follows from the singularities. (For example, there appears to be a Stokes line crossing the real  $\gamma$  axis between the values  $2/3$  and  $1$ ; if there is a flow from  $\mathcal{A}_{\text{ab}}$  to  $\mathcal{A}_-$  along this line, this might cause  $\mathcal{J}_{-,1/2}''$  to disappear from  $\mathcal{C}$ .)

The normalization of  $J_n(q)$  that is most often used in the mathematical literature differs from ours by dividing by a factor (namely the right hand side of eqn. (5.3)) that has no exponential growth in the semiclassical limit but does have a simple zero at  $\gamma = 1$ . This makes the proper statement about the semiclassical limit simpler at  $\gamma = 1$ , though not at other real values near  $\gamma = 1$ . The function  $J_n(q)$ , normalized as we do, vanishes (assuming that  $n$  is an integer) for  $q^n = 1$  or  $n = k + 2$ . One may ask what that vanishing means in the context of integration over Lefschetz thimbles. The answer is that it may be hard to see the vanishing exactly, but we can certainly see it approximately. The exponentially growing contributions cancel if  $n = k + 2$ , since after all this implies that  $k = n - 2$  is an integer. The remaining contribution is then dominated by  $\mathcal{J}_{\text{ab}}$ , but this contribution vanishes at  $n = k + 2$  since at this value, the dimension of the automorphism group of the abelian flat connection  $\mathcal{A}_{\text{ab}}$  increases.

## 6 Four Or Five-Dimensional Interpretation

In section 4.1, we observed that the flow equations of three-dimensional Chern-Simons theory have a natural four-dimensional symmetry. This is a familiar fact for the flow equations of Chern-Simons theory with a compact gauge group; what is novel and perhaps surprising is that four-dimensional symmetry also emerges in flows of complex connections. Here we will make only a few simple remarks about how this fact may be interpreted. The ultimate goal of such considerations might be to give a gauge theory interpretation to Khovanov homology, which is a four-dimensional extension of three-dimensional Chern-Simons theory.

One observation is that, in the case of a finite-dimensional Kahler manifold  $X$  with a holomorphic superpotential  $W$ , the Morse theory flow equations can be interpreted as equations for supersymmetric solitons [41]. To interpret the Chern-Simons flow equations in this way, we need a two-dimensional theory with  $\mathcal{N} = 2$  supersymmetry in which the chiral superfields are a gauge field  $\mathcal{A}$  on a three-manifold  $M$  (with some gauge group  $H$ ), the superpotential is the Chern-Simons function  $W(\mathcal{A})$ , and the gauge group is the group of maps from  $M$  to  $H$ . There is no problem in constructing such a theory. We simply start with five-dimensional maximally supersymmetric Yang-Mills theory (of course, it may be better to start with the usual ultraviolet completion of this theory in six dimensions) and compactify on  $M$  with a topological twist that preserves some supersymmetry. This gives a two-dimensional theory with  $\mathcal{N} = 2$  supersymmetry and all the stated properties. In this two-dimensional theory, the Chern-Simons flow equations that we studied in section 4.1 are equations for supersymmetric solitons. This may be a useful starting point for a new perspective on three-dimensional Chern-Simons theory.

Another perhaps more direct, though mysterious, relation between three-dimensional Chern-Simons theory and a four-dimensional theory is as follows. As we observed in section 2.6, for a given three-manifold  $M$ , the possible integration cycles of Chern-Simons theory, and therefore the possible Chern-Simons path integrals, form a vector space  $\mathcal{V}$ . For a given choice of a direction in the complex plane for the Chern-Simons coupling parameter  $k$ ,  $\mathcal{V}$  has a basis corresponding to Lefschetz thimbles associated to critical points. Because of Stokes phenomena, this basis is not really natural, but  $\mathcal{V}$  does have a natural lattice structure, since the condition for an element of  $\mathcal{V}$  to be an integer linear combination of basis vectors is invariant under all jumping processes. We can interpret  $\mathcal{V}$  as the physical “Hilbert” space of a certain topological field theory – the twisted version of  $\mathcal{N} = 4$  super Yang-Mills theory that was studied in [53] and is related to geometric Langlands. From this point of view, one studies the flow equation on a general four-manifold  $X$ , not necessarily the product of a three-manifold and a one-manifold. If  $X$  is a four-manifold of boundary  $M$ , then by counting solutions of the flow equations on  $X$ , one would construct a vector in the physical Hilbert space of  $M$  and thus an integration cycle for Chern-Simons theory on  $M$ . Of course, for this to make sense depends on the fact that the flow equations are specializations of equations

with four-dimensional symmetry.

*Acknowledgments* This research was supported in part by NSF Grant Phy-0503584. The project was stimulated by lectures at the conference *Chern-Simons Gauge Theory: 20 Years After* at the Hausdorff Institute in Bonn (August, 2009). I would like to thank the organizers for their hospitality, and to acknowledge discussions during the meeting and on other occasions with J. Anderson, C. Beasley, T. Dimofte, D. Gaiotto, S. Garoufalidis, A. Givental, M. Goresky, S. Gukov, R. MacPherson, G. Moore, and H. Murakami.

## References

- [1] W. Li, W. Song, and A. Strominger, “Chiral Gravity in Three Dimensions,” JHEP **0804**, 082 (2008), arXiv:0801.4566 [hep-th].
- [2] R. M. Kashaev, “The Hyperbolic Volume Of Knots From The Quantum Dilogarithm,” Lett. Math. Phys. **39** (1997) 269-275.
- [3] H. Murakami and J. Murakami, “The Colored Jones Polynomial And The Simplicial Volume Of A Knot,” Acta Math. **186** (2001) 85-104.
- [4] H. Murakami, J. Murakami, M. Okamoto, T. Takata, and Y. Yokota, “Kashaev’s Conjecture And The Chern-Simons Invariants Of Knots And Links,” Experiment. Math. **11** (2002) 427-435.
- [5] R. M. Kashaev and O. Tirkkonen, “A Proof Of The Volume Conjecture For Torus Knots,” J. Math. Sci. (NY) **115** (2003) 2033-2036, math.GT/9912210.
- [6] S. Gukov, “Three-Dimensional Quantum Gravity, Chern-Simons Theory, And The  $A$ -Polynomial,” Commun. Math. Phys. **255** (2005) 577-627.
- [7] H. Murakami, “Asymptotic Behaviors Of The Colored Jones Polynomials Of A Torus Knot,” Internat. J. Math. **15** (2004) 547-555.
- [8] S. Garoufalidis and T. T. Q. Lê, “An Analytic Version Of Melvin-Morton-Rozansky Conjecture,” arXiv.math.GT/0503641.
- [9] S. Garoufalidis and T. T. Q. Lê, “Asymptotics Of The Colored Jones Function Of A Knot,” arXiv:math/0508100.
- [10] J. Dubois and R. Kashaev, “On The Asymptotic Expansion Of The Colored Jones Polynomial For Torus Knots,” Math. Ann. **339** (2007) 757-782, math.GT/0510607.

- [11] S. Gukov and H. Murakami, “ $SL(2, \mathbb{C})$  Chern-Simons Theory And The Asymptotic Behavior Of The Colored Jones Polynomial,” arXiv:math/0608324.
- [12] R. van der Veen, “Proof Of The Volume Conjecture For Whitehead Chains,” arXiv:math/0611181.
- [13] H. Zheng, “Proof Of The Volume Conjecture For Whitehead Doubles Of A Family Of Torus Knots,” Chin. Ann. Math. Ser. B. **28** (2007) 375-388.
- [14] S. Garoufalidis, “Chern-Simons Theory, Analytic Continuation, And Arithmetic,” arXiv:0711.1716.
- [15] J.E Andersen & S.K. Hansen, ”Asymptotics of the Quantum Invariants of Surgeries on the Figure 8 Knot”, Journal of Knot Theory and its Ramifications **15** (2006) 1 – 69.
- [16] K. Hikami and H. Murakami, “Colored Jones Polynomials With Polynomial Growth,” arXiv:0711.2836.
- [17] H. Murakami, “An Introduction To The Volume Conjecture And Its Generalizations,” Acta Math. Vietnam **33** (2008) 209-253, arXiv:0802.0039.
- [18] T. Dimofte, S. Gukov, J. Lenells, and D. Zagier, “Exact Results For Perturbative Chern-Simons Theory With Complex Gauge Group,” arXiv:0903.2472.
- [19] R. Dijkgraaf and H. Fuji, “The Volume Conjecture And Topological Strings,” arXiv:0903.2084.
- [20] D. Zagier, “Quantum Modular Forms,” preprint.
- [21] K. Hikami and H. Murakami, “Representations And The Colored Jones Polynomial Of A Torus Knot,” arXiv:1001.2680.
- [22] D. Freed and R. Gompf, “Computer Calculations Of Witten’s 3-Manifold Invariant,” Commun. Math. Phys. **141** (1991) 79-117.
- [23] L. Jeffrey, “Chern-Simons-Witten Invariants Of Lens Spaces And Torus Bundles and the Semi-Classical Approximation,” Commun. Math. Phys. **147** (1992) 563-604.
- [24] D. Bar-Natan and S. Garoufalidis, “On The Melvin-Morton-Rozansky Conjecture,” Invent. Math. **125** (1996) 103.
- [25] L. Rozansky, “The Universal  $R$ -Matrix, Burau Representation, and the Melvin-Morton Expansion Of The Colored Jones Polynomial,” Adv. Math. **134** (1998) 1-31.
- [26] L. Rozansky, “Higher Order Terms in the Melvin-Morton Expansion of the Colored Jones Polynomial,” Commun. Math. Phys. **183** (1997) 291–306.

- [27] L. Rozansky, “A Large  $k$  Asymptotics Of Witten’s Invariant Of Seifert Manifolds,” Commun. Math. Phys. **171** (1995) 279-322, hep-th/9309099.
- [28] L. Rozansky, “The Trivial Connection Contribution to Witten’s Invariant and Finite Type Invariants of Rational Homology Spheres,” Commun. Math. Phys. **183** (1997) 23–54.
- [29] J. Andersen, “The Witten Invariant Of Finite Order Mapping Tori, I,” preprint (ca. 1995).
- [30] R. Lawrence and L. Rozansky, “Witten-Reshetikhin-Turaev Invariants Of Seifert Manifolds,” Commun. Math. Phys. **205** (1999) 287-314.
- [31] M. Mariño, “Chern-Simons Theory, Matrix Integrals, And Perturbative Three-Manifold Invariants,” Commun. Math. Phys. **253** (2004) 25-49, hep-th/0207096.
- [32] C. Beasley and E. Witten, “Non-Abelian Localization For Chern-Simons Theory,” J. Diff. Geom. **70** (2005) 183-323, hep-th/0503126.
- [33] M. Blau and G. Thompson, “Chern-Simons Theory On  $S^1$  Bundles: Abelianisation and  $q$ -Deformed Yang-Mills Theory,” JHEP 0605:003,2006, hep-th/0601068.
- [34] C. Beasley, “Localization For Wilson Loops In Chern-Simons Theory,” arXiv:0911.2687.
- [35] V. I. Arnold, S. M. Gusein-Zade, and A. N. Varchenko, “Singularities Of Differentiable Maps,” Vol. 2 (Birkhauser, 1988).
- [36] M. V. Fedorjuk, “The Asymptotics Of The Fourier Transform Of The Exponential Function Of A Polynomial,” Soviet Math. Doklady **17** (1976) 486-490.
- [37] Frédéric Pham, “Vanishing Homologies And The  $n$  Variable Saddlepoint Method,” Proc. Symp. Pure Math. **40** (1983) vol. 2, 319-333.
- [38] M. V. Berry and C. J. Howls, “Hyperasymptotics,” Proc. R. Soc. **A430** 653-668.
- [39] M. V. Berry and C. J. Howls, “Hyperasymptotics For Integrals With Saddles,” Proc. R. Soc. **A434** 657-675.
- [40] C. J. Howls, “Hyperasymptotics For Multidimensional Integrals, Exact Remainder Terms And The Global Connection Problem,” Proc. R. Soc. London **453** (1997) 2271.
- [41] S. Cecotti and C. Vafa, “On Classification of  $\mathcal{N} = 2$  Supersymmetric Theories,” Commun. Math. Phys. **158** (1993) 569 arXiv:hep-th/9211097.
- [42] K. Hori, A. Iqbal and C. Vafa, “D-Branes and Mirror Symmetry,” arXiv:hep-th/0005247.

- [43] L. Fidkowski, V. Hubeny, M. Kleban and S. Shenker, “The Black Hole Singularity in AdS/CFT,” JHEP **0402** (2004) 014, arXiv:hep-th/0306170.
- [44] J. M. Maldacena, G. W. Moore, N. Seiberg and D. Shih, “Exact vs. Semiclassical Target Space of the Minimal String,” JHEP **0410**, 020 (2004), arXiv:hep-th/0408039.
- [45] D. Gaiotto, G. W. Moore and A. Neitzke, “Four-Dimensional Wall-Crossing via Three-Dimensional Field Theory,” arXiv:0807.4723 [hep-th].
- [46] P. Kronheimer and T. Mrowka, *Monopoles And Three-Manifolds* (Cambridge University Press, 2007).
- [47] M. Hutchings, “Taubes’s Proof of the Weinstein Conjecture In Dimension Three,” Bull. Am. Math. Soc. **47** (2010) 73.
- [48] D. Bar-Natan and E. Witten, “Perturbative Expansion of Chern-Simons Theory with Noncompact Gauge Group,” Commun. Math. Phys. **141** (1991) 423.
- [49] D. Bar-Natan, “On The Vassiliev Knot Invariants,” Topology **34** (1995) 423.
- [50] S. Gukov and E. Witten, “Branes and Quantization,” arXiv:0809.0305 [hep-th].
- [51] E. Witten, “Quantum Field Theory And The Jones Polynomial,” Commun. Math. Phys. **121** (1989) 351.
- [52] K. Corlette, “Flat  $G$ -Bundles With Canonical Metrics,” J. Diff. Geom. **28** (1988) 361-382.
- [53] A. Kapustin and E. Witten, “Electric-Magnetic Duality and the Geometric Langlands Program,” Commun. Numb. Th. Phys. **1** (2007) 1-236, arXiv:hep-th/0604151.
- [54] E. Witten, “Dynamics Of Quantum Field Theory,” Lecture 7.7, in P. Deligne et. al., eds. *Quantum Fields And Strings: A Course For Mathematicians* (American Mathematical Society, 1999).
- [55] S. Gukov and E. Witten, “Gauge Theory, Ramification, and the Geometric Langlands Program,” arXiv:hep-th/0612073.
- [56] W. Nahm, “The Construction Of All Self-Dual Multimono- poles By The ADHM Method,” in *Monopoles And Quantum Field Theory*, N. S. Craigie et. al., eds. (World Scientific, 1982), 87-95.
- [57] P. Kirk and E. Klassen, “Chern-Simons Invariants Of 3-Manifolds Decomposed Along Tori And The Circle Bundle Over The Representation Space of  $T^2$ ,” Commun. Math. Phys. **153** (1993) 521.

- [58] D. Johnson and J. Millson, “Deformation Spaces Associated To Compact Hyperbolic Manifolds,” in *Discrete Groups In Geometry And Analysis: Papers In Honor of G. D. Mostow*, ed. R. Howe (Birkhauser, Boston, 1987).
- [59] E. Witten, “Chern-Simons Gauge Theory As A String Theory,” *Prog. Math.* **133** (1995) 637-678, hep-th/9207094.
- [60] S. Axelrod and I. M. Singer, “Chern-Simons Perturbation Theory, I,” in *Differential Geometric Methods In Theoretical Physics* (Proceedings, New York, 1991), hep-th/9110056.
- [61] S. Axelrod and I. M. Singer, “Chern-Simons Perturbation Theory, II,” in *J. Diff. Geom.* **39** (1994) 173-213, hep-th/9304087.
- [62] D. B. Ray and I. M. Singer, “ $R$ -Torsion And The Laplacian On Riemannian Manifolds,” *Adv. Math.* **7** (1971) 145-210.
- [63] M. F. Atiyah, “On Framings Of Three-Manifolds,” *Topology* **29** (1990) 1-7.